

# **Processo para Projetos de Ciência de Dados e Inteligência Artificial: Um estudo de caso na Secretaria Municipal de Finanças de Fortaleza/CE/Brasil**

**Ana Carla Cavalcante<sup>1</sup>, Adriano B. Albuquerque<sup>2</sup>, Vlândia Pinheiro<sup>3</sup>**

Mestrado em Computação Aplicada – Universidade Fortaleza (UNIFOR)  
Fortaleza – CE – Brazil

***Abstract.** The Data Science and Artificial Intelligence process focused on data aims to help achieve a successful science delivery, generating added value and reducing rework, as it allows a better understanding of the data and analysis of better alternatives to be applied to the problems that need to be solved. Considering the high failure rate in the delivery of this type of projects, the process presented in this paper guided the team, including engineers and data scientists, of a Brazilian Municipal Secretariat of Finance, to execute the activities needed to a successful product.*

## **1. Introdução**

A utilização de Inteligência Artificial integrada à Ciência de Dados contribui na resolução de problemas e na construção de soluções inovadoras. O que antes era tendência, hoje vem avançando rapidamente. No setor público, mesmo com menor ritmo de transformação, a aplicação dessas tecnologias propicia novas formas de otimizar processos e de colaborar para melhoria da gestão dos recursos, beneficiando os cidadãos [itforum 2023].

Gartner [apud AI & Data Today 2023], cita que 85% dos projetos que envolvem aprendizado de máquina fracassam. Além dos fracassos, vários projetos de IA não trazem o resultado pretendido e não agregam valor aos negócios, de acordo com a TechRepublic, uma organização social de profissionais de TI [apud AI & Data Today 2023].

Utilizar a qualidade de software como base para projetos de IA e Ciência de Dados, considerando diferentes guias, normas e modelos de maturidade, permite um desenvolvimento mais eficiente, confiável, escalável e com um menor índice de falhas. Mañes (2022) aborda o conjunto da *International Organization for Standardization/International Electrotechnical Commission ISO/IEC 25012:2008* – modelo geral para dados armazenados em formatos estruturados de sistemas – como um dos modelos primordiais para o desenvolvimento de produtos e serviços de qualidade dentro de sua área de atuação. Pontes (2022) complementa que a qualidade inerente aos dados se refere ao grau em que os dados satisfazem as necessidades declaradas. Na ISO/IEC 25012, o modelo apresenta características de acurácia, completude, consistência, credibilidade e atualidade.

O presente artigo descreve um processo adotado para criação de projeto de IA e Ciência de Dados aplicado à um projeto na administração tributária do município de Fortaleza (SEFIN) que tinha os seguintes objetivos principais: (1) desenvolver um modelo preditivo de valor de mercado dos imóveis considerando novas variáveis de acessibilidade e conectividade da cidade de Fortaleza, e (2) desenvolver um modelo para

descoberta de indícios de evasão fiscal do Imposto sobre Serviços. A complexidade do projeto advém dos conjuntos de dados a serem manipulados (de diversas fontes, formatos e com milhões de exemplos) e das inúmeras técnicas de Ciência de Dados e Inteligência Artificial que poderiam ser aplicadas. No início do projeto, a equipe técnica realizou uma pesquisa sobre processos de desenvolvimento que apoiassem projetos de IA e Ciência de Dados, principalmente relacionados à inovação, direcionando o time do projeto adequadamente, desde as etapas iniciais do projeto. A construção do processo aqui proposto *pari passu* a execução deste projeto na SEFIN, com todos os elementos de complexidade (tamanho, diversidade, dinamicidade e interdependência dos conjuntos de dados), foi essencial para o sucesso da entrega de um produto satisfatório ao objetivo, com a participação efetiva do cliente e, principalmente, que fosse sustentável pelo time do cliente após entrega.

A criação do processo se baseou nas melhores práticas de *Business Process Management* (BPM), na ISO 25012, na *Cognitive Project Management for Artificial Intelligence* (CPMAI) e em abordagens ágeis.

## **2. Processo e Estudo de Caso**

No processo foram definidas 4 fases: (i) Planejamento do projeto, (ii) Coleta de dados; (iii) Desenho e treinamento do Modelo e (iv) Implantação, Manutenção e Capacitação. A fase Planejamento do projeto contempla as atividades: 1. Compreender o desafio, 2. Analisar documentações, 3. Criar *Canvas Statement Problem*, 4. Criar *Software Analytics Canvas*, 5. Especificar requisitos de qualidade de dados a partir da ISO 25012, 6. Definir abordagem de evolução do projeto, 7. Definir plano técnico, 8. Elaborar plano de projeto, 9. Apresentar aos *Stakeholders* o plano de projeto.

Os links para a representação gráfica do processo, para o *Problem Statement Canvas*, para o *System Analytics – Dados* e para o Resultado da Pesquisa de Avaliação do Processo estão disponibilizados respectivamente nos Apêndices A, B, C e D.

A fase Coleta de Dados abrange as seguintes atividades: 10. Selecionar problemas elencados, 11. Coletar dados brutos, 12. Definir os dados que serão trabalhados, 13. Realizar a limpeza dos dados, 14. Normalizar os dados, 15. Realizar análise exploratória dos dados, 16. Gerar os *datasets*, 17. Apresentar gráficos dos dados analisados.

No caso da fase Desenho e Treinamento do Modelo, fazem parte as atividades: 18. Definir técnicas de ML/Algoritmos/Técnicas que serão aplicadas aos dados, 19. Selecionar features que serão utilizadas no treinamento do modelo, 20. Aplicar algoritmos preditivos aplicados aos dados, 21. Treinar modelos, 22. Visualizar o treinamento dos algoritmos, 23. Validar modelos e 24. Elaborar representação visual dos dados, caso o modelo tenha sido validado.

A fase Implantação, Manutenção e Capacitação abrange: 25. Apresentar produto ao cliente, 26. Realizar a implantação do produto, 27. Realizar manutenções necessárias, 28. Capacitar os usuários e 29. Elaborar documento consolidado do projeto.

As atividades 1 a 9 estão relacionadas à fase de planejamento do projeto e foram de grande relevância. Os principais artefatos de saída dessas atividades foram: *Statement*

*Problem Canvas*, *Software Analytics Canvas* e o Plano do Projeto. Este último incluía a equipe envolvida, funções e responsabilidades, cronograma de ações/entregas, ciclos de *sprint* e as respectivas ferramentas colaborativas que deveriam ser adotadas, requisitos técnicos necessários para o desenvolvimento do trabalho e pontos de controle para apresentar a evolução do projeto às partes interessadas.

As atividades 10 a 17 estão relacionadas à fase de coleta de dados e foram realizadas pela equipe de tecnologia, neste caso, cientistas de dados e engenheiros. O grande desafio em projetos de ciência de dados não é apenas a disponibilidade de dados, mas também a qualidade dos dados. Dados de baixa qualidade prolongam o tempo do projeto, além de criar modelos que não aderem à solução de problemas. Portanto, nessa etapa, foi necessário garantir não só a alta qualidade dos dados, mas também que fossem relevantes para o problema que seria abordado.

As atividades 18 a 24 estão relacionadas à fase de desenho e ao treinamento do modelo escolhido. A definição da técnica a ser utilizada em um projeto de *Data Science* dependeu de um conjunto de fatores, como: o problema a ser abordado, o tipo de resultado que se desejava ser alcançado e o tamanho dos dados, por exemplo.

As atividades 25 a 29 estão relacionadas à fase de implantação, manutenção e capacitação. Nessa fase, os resultados finais e consolidados foram apresentados às partes interessadas, lembrando que, à medida que o projeto evoluiu, nas reuniões planejadas, os resultados obtidos no decorrer do projeto foram compartilhados para que, se necessário, a rota pudesse ser ajustada. Então, após essa apresentação final, seguiu-se para a implementação do produto. Todas as funcionalidades oferecidas pelo produto foram conhecidas pelos usuários finais, uma vez que seu uso correto proporciona ao cliente o valor agregado esperado. A última atividade referiu-se à entrega de um documento consolidado do projeto, fornecendo uma documentação completa das técnicas e algoritmos de aprendizado de máquina utilizados no projeto, orientando como realizar a análise adequada dos artefatos de saída e sua manutenção.

### **3. Conclusão**

A construção desse processo permitiu ajudar a tratar alguns problemas existentes em projetos que envolvem Ciência de Dados e Inteligência Artificial, como:

- Direcionamento de um time, principalmente cientistas de dados e engenheiro de dados, para compreender realmente as dores que o negócio buscava resolver, participando da construção de forma efetiva do *Statement Problem Canvas* e *Software Analytics Canvas*;
- Cientistas de dados e Engenheiro de dados validaram as atividades centrais das fases de Coleta de Dados e Desenho e Treinamento do Modelo;
- Sustentabilidade de um projeto de Ciência de Dados e IA independente do time que o implementou. No estudo de caso do artigo, a sustentabilidade está sendo executada pelo time do cliente, contemplando as atividades constantes na fase Implantação, Manutenção e Capacitação.

O processo construído é uma abordagem centrada em dados de negócios, com a participação ativa das partes interessadas. Além de ter dado indícios de ser eficaz,

garantiu uma melhor qualidade dos dados, melhorou a comunicação entre os membros da equipe, garantiu uma maior sustentabilidade e agregou valor ao cliente.

## Referências

itforum.(2023) "Transformação digital do setor público esbarra em alta administração"

<https://itforum.com.br/noticias/transformacao-digital-do-setor-publico-esbarra-em-alta-administracao/>, Março.

AI & Data Today.(2023) "Top 10 Reasons Why AI Projects Fail"

<https://www.aidatoday.com/top-10-reasons-why-ai-projects-fail/#:~:text=The%20rate%20of%20AI%20project,intended%20results%20to%20the%20business,> Março.

Mañes, Juan.(2022) "Technical Standards to achieve Data Quality"

<https://datos.gob.es/en/blog/technical-standards-achieve-data-quality>, Maio.

Pontes, Leonardo Bastos.(2022) "BIDS – A Business Intelligence and Data Science Development Process Supported by Principles of Design Thinking, ISO/IEC 25000, and Rational Unified Process." Tese (Doutorado) – Universidade de Fortaleza.

## Apêndices

### Apêndice A - Processo

Link: [Processo para Projetos de Ciência de Dados e Inteligência Artificial](#)

### Apêndice B - *Template Problem Statement Canvas*

Link: [Template Problem Statement Canvas](#)

### Apêndice C - *Template System Analytics - Dados*

Link: [Template System Analytics - Dados](#)

### Apêndice D - Resultado da Pesquisa de Avaliação do Processo

Link: [Resultado da Pesquisa de Avaliação do Processo](#)