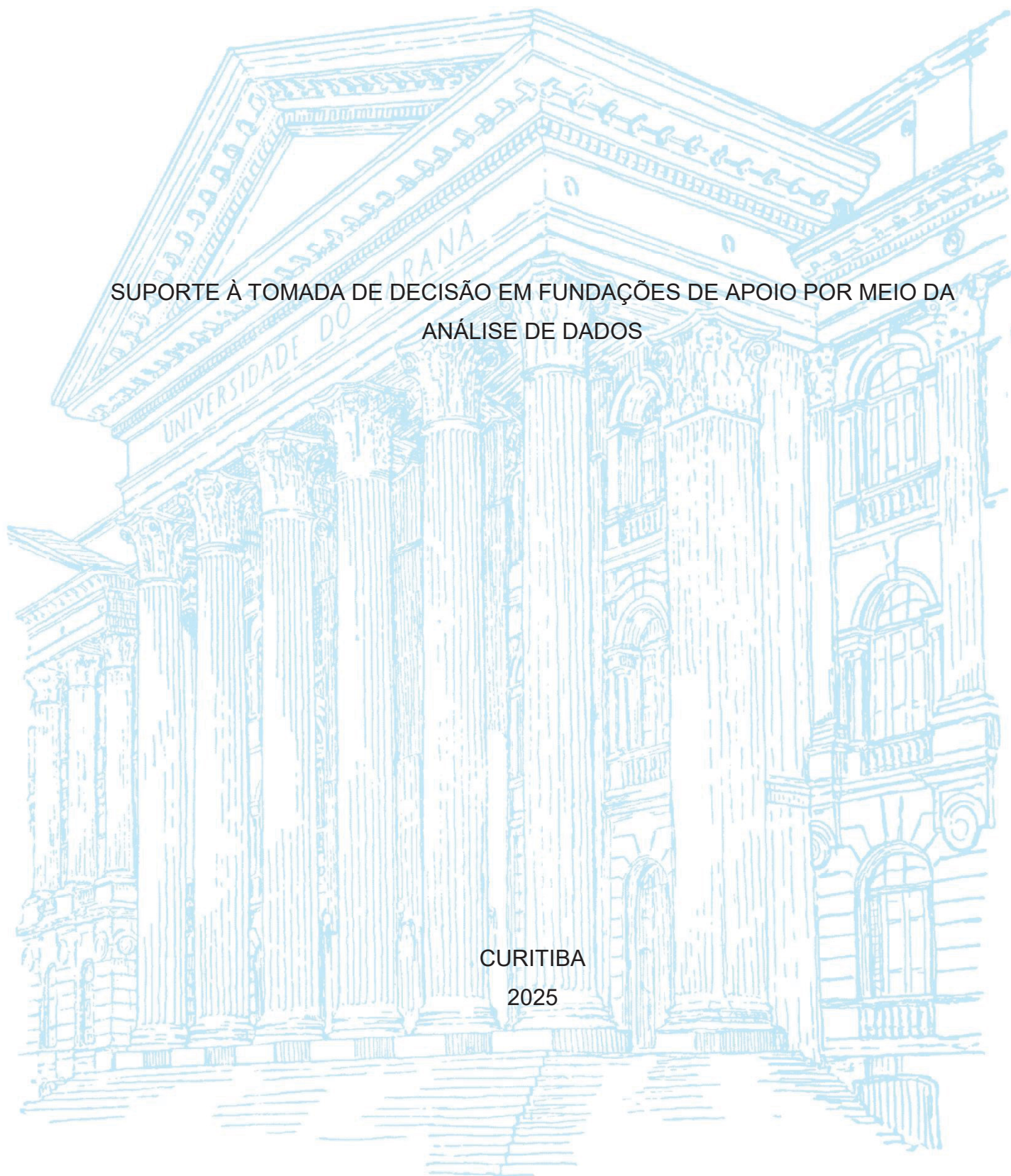


UNIVERSIDADE FEDERAL DO PARANÁ

MARCOS CESAR MIRANDA DA SILVA

SUPOORTE À TOMADA DE DECISÃO EM FUNDAÇÕES DE APOIO POR MEIO DA
ANÁLISE DE DADOS

CURITIBA
2025



MARCOS CESAR MIRANDA DA SILVA

SUPORTE À TOMADA DE DECISÃO EM FUNDAÇÕES DE APOIO POR MEIO DA
ANÁLISE DE DADOS

Dissertação apresentada ao Curso de Pós-Graduação em Gestão da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Gestão da Informação.

Orientador: Prof. Dr. José Marcelo Almeida Prado Cestari

Coorientadora: Profa. Dra. Denise Fukumi Tsunoda

CURITIBA

2025

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

UNIVERSIDADE FEDERAL DO PARANÁ

SISTEMA DE BIBLIOTECAS – BIBLIOTECA CIÊNCIAS SOCIAIS APLICADAS

Silva, Marcos Cesar Miranda da

Suporte à tomada de decisão em Fundações de Apoio por meio da análise de dados / Marcos Cesar Miranda da Silva. – Curitiba, 2025. 1 recurso on-line : PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Gestão da Informação.

Orientador: Prof. Dr. José Marcelo Almeida Prado Cestari.
Coorientadora: Profa. Dra. Denise Fukumi Tsunoda.

1. Gestão da Informação. 2. Análise de dados. 3. Mineração de dados (Computação). 4. Tomada de Decisão. I. Cestari, José Marcelo Almeida Prado. II. Tsunoda, Denise Fukumi. III. Universidade Federal do Paraná. Programa de Pós-Graduação em Gestão da Informação. IV. Título.

Bibliotecário: Nilson Carlos Vieira Junior - CRB-9/1797



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS SOCIAIS E APLICADAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GESTÃO DA
INFORMAÇÃO - 4001016058P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação GESTÃO DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a avaliação da Dissertação de Mestrado de **MARCOS CESAR MIRANDA DA SILVA** (Iniciais: **SUPORTE À TOMADA DE DECISÃO EM FUNDAÇÕES DE APOIO POR MEIO DA ANÁLISE DE DADOS**), sob orientação do Prof. Dr. **JOSE MARCELO ALMEIDA PRADO COSTARI**, que após terem lido o texto e realizado a avaliação do trabalho, soube parecer pela sua **Aprovação** no ato de defesa.
A outorga do título de Mestre está sujeita a homologação pelo Colegiado, ao atendimento de todas as condições e prescrições estabelecidas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 33 de Outubro de 2020.


JOSE MARCELO ALMEIDA PRADO COSTARI
Presidente da Banca Examinadora


EDUARDO DE FREITAS ROCHA LOURES
Avaliador Externo (PONTIFÍCA UNIVERSIDADE CATÓLICA DO PARANÁ)


EDUARDO ALVES PORTELA SANTOS
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ - DEPARTAMENTO DE ADMINISTRAÇÃO GERAL E APLICADA)

Dedico este trabalho, primeiramente a Deus, à minha esposa e ao meu filho.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus pela dádiva da vida, pela força nos momentos difíceis e pela luz nos caminhos percorridos.

À minha esposa Nadia Hinckel de Freitas, e ao meu filho Lucca de Freitas da Silva, meus pilares, por todo amor, paciência e apoio incondicional durante essa intensa jornada.

Expresso minha profunda gratidão ao meu orientador, Prof. Dr. José Marcelo Prado de Almeida Cestari, e a minha coorientadora, Profa. Denise Fukumi Tsunoda, por aceitarem orientar este trabalho, pela orientação atenciosa, pelo incentivo constante e pela confiança depositada em mim.

Sou especialmente grato pelas valiosas discussões e reuniões, que contribuíram significativamente para o desenvolvimento desse projeto e para o meu crescimento acadêmico e pessoal.

Agradeço também aos membros da banca Prof. Dr. Eduardo Alves Portela Santos e Prof. Dr. Eduardo de Freitas Rocha Loures, pelas contribuições e apontamentos que foram de extrema importância para o aprimoramento desta pesquisa.

Aos professores e colegas do Programa de Pós-Graduação em Gestão da Informação da Universidade Federal do Paraná, meu reconhecimento pela convivência enriquecedora ao longo desta trajetória.

Agradeço, ainda, à empresa onde atuo, por compreender a importância desta etapa e permitir a flexibilidade necessária para que eu pudesse participar das aulas e cumprir as exigências deste mestrado.

A todos, o meu mais sincero obrigado.

RESUMO

Esta dissertação tem como objetivo analisar o potencial da aplicação de técnicas de análise exploratória e mineração de dados na identificação de padrões, tendências e oportunidades de melhoria nos processos operacionais e de gestão de uma fundação de apoio, visando subsidiar a tomada de decisão estratégica e a sustentabilidade institucional vinculada a uma Instituição Federal de Ensino Superior. O estudo utilizou dados operacionais de projetos geridos por uma fundação de apoio à Universidade Federal do Paraná (UFPR) no período de 2019 a 2024, coletados por meio do sistema ERP Protheus. A metodologia adotada foi baseada no modelo CRISP-DM, que orientou o processo em seis etapas: compreensão do negócio, compreensão dos dados, preparação, modelagem, avaliação e implantação. Após o tratamento da base de dados (com 270 projetos válidos), foram aplicadas técnicas de classificação (*Random Forest*), regressão, clusterização (*k-means*) e regras de associação (*Apriori*). Os resultados revelaram que apenas 8,88% dos projetos apresentaram recuperação de custos operacionais (RCO) igual ou superior ao calculado, e cerca de 78% apresentaram desempenho inferior a 70%, sendo as principais limitações atribuídas às normativas legais que impõem tetos para a cobrança das Despesas Operacionais Administrativas (DOA). O modelo de regressão para o RCO efetivo obteve erro médio de 6,5%, e a clusterização evidenciou três perfis distintos de projetos quanto a rentabilidade e risco. A partir dessas análises, foi proposto um *framework* de governança de dados e apoio à gestão institucional, com foco na padronização do cadastro de projetos, criação da figura do *data steward*, e uso contínuo de painéis analíticos como suporte à tomada de decisão. A pesquisa contribui para o fortalecimento da gestão estratégica da informação em fundações de apoio, propondo ações práticas para melhoria da sustentabilidade financeira, captação de recursos e eficiência nos processos internos.

Palavras-chave: Análise de dados, Mineração de dados, CRISP-DM, Fundações de apoio, Recuperação de custos operacionais.

ABSTRACT

This dissertation aims to analyze the potential of applying exploratory data analysis and data mining to identify patterns, trends, and opportunities for improvement in the operational and management processes at a supporting foundation, with a view to supporting strategic decision-making and institutional linked to a Federal Higher Education Institution. The study used operational data from projects managed by a foundation supporting the Federal University of Paraná (UFPR) from 2019 to 2024, collected through the Protheus ERP system. The methodology adopted was based on the CRISP-DM model, which guided the process in six stages: understanding the business, understanding the data, preparation, modeling, evaluation and implementation. After processing the database (with 270 valid projects), classification (Random Forest), regression, clustering (k-means) and association rules (Apriori) techniques were applied. The results showed that only 8.88% of the projects had an operating cost recovery (RCO) equal to or greater than that initially calculated, and around 78% had a performance of less than 70%, the main limitations being attributed to legal regulations that impose ceilings on the collection of DOA (Operating Administrative Costs). The regression model for the effective RCO obtained an average error of 6.5%, and the clustering showed three distinct project profiles in terms of profitability and risk. Based on these analyses, a data governance and institutional management support framework was proposed, focusing on standardizing project registration, creating the role of data steward, and continuously using analytical dashboards to support decision-making. The research contributes to strengthening strategic information management in support foundations, proposing practical actions to improve financial sustainability, fundraising and efficiency of internal processes.

Keywords: Data analysis, Data mining, CRISP-DM, Support foundations, Recovery of operational cost.

LISTA DE FIGURAS

FIGURA 1 – FASES DO PROCESSO KDD.....	28
FIGURA 2 – PASSOS DA KDD.....	29
FIGURA 3 - SÍNTESE DAS ETAPAS DE PESQUISA.....	50
FIGURA 4 - ETAPAS DA ANÁLISE EXPLORATÓRIA DE DADOS (AED)	52
FIGURA 5 – CLASSE RENTABILIDADE	75
FIGURA 6 – <i>CLUSTERS</i> DE PROJETOS RCO EFETIVO x RECEITA REALIZADA	79
FIGURA 7 – <i>FRAMEWORK</i> PROPOSTO PARA APOIO À GESTÃO DE PROJETOS	92

LISTA DE QUADROS

QUADRO 1 – ABORDAGENS ESTATÍSTICAS E ESTRATÉGIAS UTILIZADAS	26
QUADRO 2 – FLUXO DE GERAÇÃO DE VALOR: DA BASE DE DADOS AO SUPORTE À DECISÃO	35
QUADRO 3 – PANORAMA DA ESTRATÉGIA DE BUSCA	43
QUADRO 4 – ARTIGOS SELECIONADOS	44
QUADRO 5 – LEVANTAMENTO BIBLIOGRÁFICO COMPLEMENTAR	46
QUADRO 6 - ETAPAS CRISP-DM.....	51
QUADRO 7 – RELAÇÃO ENTRE RCO EFETIVO X RCO CALCULADO (em milhares de reais)	65
QUADRO 8 – RELAÇÃO ENTRE O RECURSO PREVISTO X RECURSO EFETIVO (em milhares de reais)	67
QUADRO 9 – RENTABILIDADE POR TIPO.	69
QUADRO 10 – COMPARATIVO DE ALGORITMOS APLICADOS.	72
QUADRO 11 – RESULTADO TESTE <i>RANDOM FOREST</i>	73
QUADRO 12 – MATRIZ DE CONFUSÃO E MÉTRICAS DETALHADAS POR CLASSE (<i>RANDOM FOREST</i>)	75
QUADRO 13 – COEFICIENTE DE SILHUETA MÉDIO POR NÚMERO DE <i>CLUSTERS</i> (<i>k</i>).....	78
QUADRO 14 – ESTATÍSTICAS POR <i>CLUSTER</i> X TICKET MÉDIO X RECEITA X QUANTIDADE DE PROJETOS	80
QUADRO 15 – DISTRIBUIÇÃO DAS CLASSES DE RENTABILIDADE X <i>CLUSTER</i>	81
QUADRO 16 – PONTOS FORTES, FRACOS E OPORTUNIDADES X <i>CLUSTERS</i>	82

LISTA DE ABREVIATURAS E SIGLAS

AED	- Análise Exploratória de Dados
ANDIFES	- Associação Nacional das Instituições Federais de Ensino Superior
<i>*ba*</i>	- Lugar ou espaço do conhecimento
CGU	- Controladoria Geral da União
CONFIES	- Conselho Nacional das Fundações de Apoio às Instituições de Ensino Superior e de Pesquisa Científica e Tecnológica
DOA	- Despesas Operacionais Administrativas
e-MEC	- Sistema de Regulação do Ensino Superior
GI	- Gestão da Informação
GC	- Gestão do Conhecimento
IBICT	- Instituto Brasileiro de Informação em Ciência e Tecnologia
ICTs	- Instituições Científicas, Tecnológicas e de Inovação
IFES	- Instituição Federal de Ensino Superior
<i>Informis</i>	- <i>Institute for Operations Research and Management Sciences</i>
KDD	- Descoberta de conhecimento em banco de dados
<i>MAE</i>	- <i>Mean Absolute Error</i>
<i>MSE</i>	- <i>Mean Squared Error</i>
MEC	- Ministério da Educação
RCO	- Recuperação de Custos Operacionais
<i>RMSE</i>	- <i>Root Mean Squared Error</i>
SECI	- Socialização, Externalização, Combinação e Internalização
SICONV	- Sistema de Convênios e Contratos de Repasse
TCU	- Tribunal de Contas da União
UFPR	- Universidade Federal do Paraná

SUMÁRIO

1 INTRODUÇÃO	13
1.1 PROBLEMA DE PESQUISA	15
1.2 OBJETIVOS	17
1.2.1 Objetivo geral	17
1.2.2 Objetivos específicos.....	17
1.3 JUSTIFICATIVA	17
2 REFERENCIAL TEÓRICO.....	20
2.1 FUNDAÇÃO PRIVADA.....	20
2.1.1 Fundação de Apoio	22
2.2 ANÁLISE EXPLORATÓRIA DE DADOS (AED).....	25
2.3 MINERAÇÃO DE DADOS	26
2.4 TOMADA DE DECISÃO	31
2.5 GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO	35
3 TRAJETÓRIA METODOLÓGICA.....	41
3.1 MÉTODO.....	41
3.2 LEVANTAMENTO BIBLIOGRÁFICO	42
4 PROPOSTA METODOLÓGICA.....	50
4.1 MODELO DE REFERÊNCIA PARA O PROCESSO ANÁLITICO: CRISP-DM....	50
4.2 ETAPAS DA ANÁLISE EXPLORATÓRIA DE DADOS (AED).....	52
5 APLICAÇÕES.....	56
5.1 COMPREENSÃO DO NEGÓCIO	56
5.2 COMPREENSÃO DOS DADOS.....	58
5.3 PREPARAÇÃO DOS DADOS	60
5.4 MODELAGEM.....	64
5.4.1 Aplicação da Análise Exploratória de Dados.....	65
5.4.2 Aplicação da Mineração de Dados.....	71
5.5 AVALIAÇÃO	82
5.6 IMPLANTAÇÃO.....	85
6 RESULTADOS E ANÁLISES	87
6.1 ANÁLISE EXPLORATÓRIA DA BASE DE PROJETOS.....	87
6.2 APLICAÇÃO DA METODOLOGIA <i>CRISP-DM</i> PROJETOS	88
6.3 RESULTADO DAS ANÁLISES.....	88

6.4 RENTABILIDADE POR TIPO DE PROJETO	89
6.5 APLICAÇÃO DOS ALGORITMOS DE MINERAÇÃO DE DADOS	89
6.6 PROPOSTA DE <i>FRAMEWORK</i> PARA GOVERNANÇA DE DADOS E APOIO À GESTÃO DE PROJETOS	90
6.6.1 Passo a passo para utilização do Framework	93
6.7 CONSIDERAÇÕES FINAIS DA ANÁLISE	94
6.7.1 Contribuições Científicas e Acadêmicas	95
6.7.2 Alinhamento com a Linha de Pesquisa	96
7 CONSIDERAÇÕES FINAIS	99
7.1 ATENDIMENTO AOS OBJETIVOS.....	100
7.1.1 Objetivos específicos.....	101
7.2 SUGESTÕES PARA TRABALHOS FUTUROS	101
REFERÊNCIAS.....	104

1 INTRODUÇÃO

A transformação de dados operacionais em informações e conhecimento é essencial para subsidiar a tomada de decisão estratégica nas fundações de apoio, que operam sob forte pressão por transparência e sustentabilidade.

Apesar do volume de dados disponíveis, há uma lacuna entre a coleta e o uso efetivo dessas informações, o que compromete decisões operacionais, táticas e estratégicas.

Esta pesquisa apresenta as fundações de apoio, instituições criadas por lei para apoiar e dar suporte aos projetos de ensino, pesquisa, extensão, desenvolvimento institucional e tecnológico e inovação para Instituições Federais de Ensino Superior (IFES) e Instituições Científicas, Tecnológicas e de Inovação (ICTs).

Apesar de sua característica de apoio a uma instituição pública, é uma fundação de direito privado e está inserida em ambientes cada vez mais dinâmicos e exigentes.

Precisando enfrentar desafios como otimizar os seus procedimentos operacionais, administrativos e estratégicos sem deixar de lado a crescente demanda da sociedade e dos órgãos de controle por eficiência, transparência e resultados melhores.

As fundações desempenham um papel fundamental na captação e gestão de recursos voltados para projetos de ensino, pesquisa, extensão, inovação e desenvolvimento institucional e científico-tecnológico, atuando como elos entre a academia, o mercado e a sociedade (CONFIES, 2024).

No entanto, desafios como a escassez de recursos e a complexidade dos processos internos exigem abordagens inovadoras para otimizar a gestão e fortalecer a sustentabilidade dessas instituições.

Nesse contexto, a produção e o armazenamento de grandes volumes de dados operacionais gerados pela fundação tornam-se elementos centrais para a gestão, especialmente com o uso de sistemas integrados como o ERP Protheus da TOTVS.

O sistema de gestão atualmente utilizado pela fundação apresenta um grande potencial em volume e características de dados que poderiam ser usados para a tomada de decisão, no entanto, a sua utilização é limitada e passiva.

A ausência de uma análise de dados estruturada acaba impedindo que as informações que estão registradas nos sistemas internos sejam convertidas em conhecimento útil para a definição de metas e estratégias institucionais. Seja pela falta de entendimento da informação ou pela falta de normalização e classificação correta dos dados.

A principal consequência para a fundação é que oportunidades de melhoria nos procedimentos internos podem ser negligenciadas, seja na gestão de contratos, na captação de recursos ou na gestão e execução do projeto como um todo.

A fundação tem trabalhado em algumas iniciativas pontuais, com a integração do sistema de gestão e utilização de *dashboards* na plataforma de *Power BI*, para o controle e acompanhamento da execução dos projetos que são realizados.

No entanto, o principal gargalo é que essas ações não estão fundamentadas em metodologias de análise de dados, e nem contam com uma abordagem sistemática de extração do conhecimento.

As técnicas de análise de dados permitem identificar e extrair percepções que contribuem para decisões mais assertivas, auxiliando na mitigação de riscos e na resolução de problemas, além de revelar novas oportunidades de negócios.

Ademais, essas ferramentas oferecem múltiplas opções estratégicas, desempenhando um papel fundamental no apoio à tomada de decisão em todos os níveis hierárquicos, do operacional ao estratégico.

A situação atual da fundação pesquisada configura um campo fértil para o uso de ferramentas analíticas, como a análise exploratória e a mineração de dados para que, através desse conhecimento, se possa transformar dados operacionais em informações estratégicas.

A gestão de dados tem se tornado um fator estratégico dentro das organizações e são base para a tomada de decisão dessa forma se encaixam e podem ser aplicados nesse estudo. Huang e Yue (2024) destaca que a análise de dados desempenha um papel importante ao transformar informações brutas em *insights* que auxiliem nas decisões.

A AED pode ser uma ferramenta muito importante para as organizações a fim de possibilitar a obtenção de informações de forma útil, cada vez mais rápida e consolidada.

Diante disso, trazem-se percepções e caminhos a seguir, muitas vezes ocultos nos dados, o que não seria possível sem sua aplicação. Os dados são os

insumos da informação e do conhecimento e por consequência a chave para a análise e tomada de decisão nas organizações.

De acordo com Marques e Correia Neto (2020) no processo de tomada de decisão, é essencial que a decisão seja bem fundamentada, especialmente em situações que possam trazer riscos à organização.

Nesse sentido, por exemplo, uma fundação poderia aceitar fazer a gestão de um projeto sem previsão de repasses financeiros bem definidos ou com cronograma de execução inexecutável.

Para evitar riscos à organização o gestor deve garantir que sua escolha seja amparada por um contexto claro e bem definido, já que decisões mal tomadas podem resultar em graves consequências, incluindo problemas legais. Tomar decisões é uma atividade constante no ambiente organizacional, onde o foco está em buscar soluções que agreguem valor às partes interessadas.

A qualidade das decisões está diretamente relacionada à eficácia dos processos que conduzem a esses resultados, sendo importante garantir rapidez e precisão nos julgamentos para alcançar decisões de alta qualidade (Abu-Alsodos, 2023).

No contexto organizacional, é necessário tomar decisões cotidianas acerca dos rumos institucionais, demandando suporte informacional, cuja ausência pode comprometer a assertividade das escolhas e acarretar prejuízo a organização.

A pesquisa exploratória será realizada em uma fundação que gera diariamente uma grande quantidade de dados em seus processos internos, desde as operações mais básicas até as mais estratégicas.

Esses dados gerados podem ser verificados na apresentação anual do seu relatório de atividades. A pesquisa tem por objetivo em realizar uma análise de dados operacionais de uma fundação de apoio vinculada à Universidade Federal do Paraná (UFPR).

1.1 PROBLEMA DE PESQUISA

Na presente pesquisa, apresentam-se as teorias de análise exploratória e mineração de dados como ferramentas de apoio à tomada de decisão, desempenhando um papel relevante na descoberta de conhecimento por meio da análise, compreensão e correlação entre variáveis.

Esse projeto propõe o desenvolvimento de um *framework* de extração de conhecimento baseado em técnicas de mineração de dados, com foco no suporte à tomada de decisão estratégica no contexto de uma Fundação de Apoio.

O objetivo principal consiste em analisar os dados provenientes dos sistemas operacionais da fundação para identificar variáveis-chave, padrões de comportamento e relações significativas que possam orientar ações gerenciais mais eficazes, especialmente no setor estratégico, de captação e gestão recursos, de gestão de projetos e contratos.

A proposta está ancorada em uma abordagem exploratória e aplicada, com potencial para gerar impactos práticos significativos na cultura organizacional e nos mecanismos de governança da fundação, além de contribuir para o avanço da pesquisa na área da ciência de dados aplicada à gestão de Fundações de Apoio às Instituições Federais de Ensino Superior (IFES) e Instituições de Ciência de Tecnologia (ICTs).

Em síntese, o problema de pesquisa está relacionado à dificuldade enfrentada pela Fundação de Apoio em transformar o grande volume de dados operacionais disponíveis em informações e conhecimentos estratégicos capazes de subsidiar a tomada de decisão institucional.

Essa limitação compromete a capacidade da fundação em superar desafios orçamentários, ampliar a captação de recursos e, por consequência, aumentar o valor do ressarcimento das Despesas Operacionais Administrativas (DOA).

O problema de pesquisa reside na dificuldade de transformar dados operacionais em insumos para a decisão estratégica, impactando a capacidade da fundação em identificar gargalos, priorizar projetos rentáveis e mitigar riscos financeiros.

Assim, pergunta-se **de que forma a aplicação estruturada de técnicas de análise exploratória e mineração de dados podem apoiar o processo de tomada de decisão nas fundações de apoio?**

A premissa da pesquisa parte do entendimento de que a aplicação estruturada dessas técnicas pode permitir a extração do conhecimento estratégico a partir dos dados operacionais da fundação, fortalecendo o processo de tomada de decisão e promovendo maior eficiência na gestão institucional.

1.2 OBJETIVOS

No intuito de responder ao problema de pesquisa foram elaborados o objetivo geral e os objetivos específicos, conforme descritos nos tópicos 1.2.1 e 1.2.2.

1.2.1 Objetivo geral

O objetivo geral da pesquisa é analisar o potencial da aplicação de técnicas de análise exploratória e mineração de dados na identificação de padrões, tendências e oportunidades de melhoria nos processos operacionais e de gestão de uma fundação de apoio que subsidiem ações gerenciais e processos decisórios.

1.2.2 Objetivos específicos

Com o intuito de responder ao objetivo geral foram definidos os objetivos específicos da pesquisa, conforme descrição abaixo:

- a) analisar os dados disponíveis na base de dados da fundação para identificar variáveis-chave que influenciam na gestão e operação de projetos;
- b) identificar padrões e relações nos dados que possam indicar oportunidades de melhoria na gestão institucional; e
- c) estabelecer recomendações baseadas nos dados analisados, desenvolver e apresentar um *framework* de governança de dados, otimizar os processos internos e a tomada de decisão estratégica com foco na captação de recursos.

1.3 JUSTIFICATIVA

Segundo CONFIES (2024), as fundações de apoio foram criadas com o propósito de fortalecer as Instituições Federais de Ensino Superior (IFES) e as Instituições Científica, Tecnológica e de Inovação (ICTs), desempenhando um papel importante na promoção do crescimento da ciência, da pesquisa e do desenvolvimento tecnológico no Brasil.

Por meio de uma gestão eficiente e estratégica, essas entidades otimizam a captação e a aplicação de recursos financeiros, beneficiando diretamente a comunidade universitária e a sociedade como um todo.

De acordo com Abreu (2024), as fundações não se limitam à administração de recursos, mas também fomentam a inovação e asseguram a transparência e legalidade nos processos administrativos.

Com esse suporte, permitem que as IFES e ICTs concentrem esforços em suas atividades-fim, como produção de conhecimento, a formação de profissionais qualificados e a contribuição para o desenvolvimento social e econômico do país.

Em consulta ao portal eletrônico do Ministério da Educação (MEC), o e-MEC, e da Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior (ANDIFES), identificou-se que o Brasil conta atualmente com 69 IFES.

Além disso, conforme relatório disponível no portal do MEC, existem 106 fundações de apoio credenciadas ou autorizadas para prestar suporte na gestão administrativa e financeira de projetos para essas instituições.

Segundo o CONFIES (2024), mesmo com a redução do orçamento público para ciência e tecnologia, as fundações de apoio de universidades públicas e institutos de pesquisa registraram, em 2021, receita superior a R\$ 8 bilhões, recorde na captação de recursos para a ciência no Brasil.

Esses dados, obtidos a partir de uma pesquisa realizada com 94 instituições afiliadas, das quais 50 participaram efetivamente, destacam o papel fundamental das fundações na sustentação e avanço da pesquisa científica no país.

Os cortes e os contingenciamentos orçamentários e financeiros têm influenciado a qualidade do ensino, suspensão de obras, manutenções e serviços essenciais das universidades (Caetano; Campos, 2019).

As fundações de apoio foram criadas com o intuito de ajudar as universidades federais a superarem essas dificuldades, atuando como um elo entre a universidade e a sociedade, e entre a universidade e o mercado.

Nesse contexto, as fundações atuam em todo o processo, desde a captação de recursos e submissão de propostas até a formalização de contratos, gestão administrativa e financeira e prestação de contas.

As fundações permitem que os pesquisadores se concentrem exclusivamente em suas pesquisas e não tenham que dedicar tempo com as questões administrativas e burocracias necessárias para o andamento dos projetos.

Contudo, com a crescente escassez de recursos e orçamentos limitados, surge a necessidade de não perder o tempo da pesquisa e do desenvolvimento tecnológico (Caetano; Campos, 2019). Esse é o desafio que direciona a busca por soluções para o problema de pesquisa.

Como as fundações não possuem participação no orçamento público (pois são de direito privado) e sua função é a gestão administrativa e financeira dos projetos, sem contar com pesquisadores próprios, precisam criar estratégias eficazes para captar mais projetos e garantir sua sustentabilidade.

Assim, as fundações podem cobrir seus custos administrativos, operacionais e financeiros, mantendo sua sustentabilidade e garantindo o apoio contínuo à Instituição Apoiada e à sociedade paranaense.

A fundação de apoio, objeto da pesquisa exploratória, apresenta como objetivos principais, descritos em seu estatuto:

“Promover e apoiar programas e eventos relacionados ao ensino, pesquisa, extensão, desenvolvimento institucional, científico e tecnológico, assistência social, saúde, cultura, proteção ambiental e outras atividades correlatas, alinhadas às políticas da Universidade Federal do Paraná (UFPR) (FUNDAÇÃO, 2018)”.

Ainda busca viabilizar recursos técnicos, gerenciais e financeiros para atividades vinculadas à UFPR e instituições conveniadas, além de conceder bolsas, colaborar com diversas entidades e desenvolver atividades de consultoria.

Para tanto, estabelece acordos e contratos, podendo ainda criar filiais e escritórios no país e no exterior. Com mais de 45 anos de atuação, consolidou-se como apoio essencial à gestão de projetos vinculados à UFPR.

Em 2023, foram geridos 468 projetos, envolvendo 51.334 protocolos, mais de 99 mil conciliações bancárias, 185 mil movimentações financeiras, 2.953 processos de aquisição e 1.435 empregados CLT, totalizando R\$ 215,5 milhões em operações.

Esses dados evidenciam a complexidade de sua atuação e reforçam a necessidade de métodos e ferramentas capazes de transformar dados em informações estratégicas para subsidiar a tomada de decisão.

A análise eficiente desses dados pode contribuir significativamente para otimizar os processos internos e ampliar o impacto das ações realizadas em apoio à UFPR e à sociedade.

2 REFERENCIAL TEÓRICO

Neste capítulo, a fim de trazer um aprofundamento teórico são apresentados os estudos referentes as fundações, a análise exploratória de dados, mineração de dados, tomada de decisão e gestão da informação.

2.1 FUNDAÇÃO PRIVADA

Neste tópico são apresentados os principais conceitos relacionados a esse tipo de organização, objeto desse estudo.

Segundo Paes (2017), ela se diferencia tanto de empresas privadas quanto de empresas públicas, de um lado ficava o Estado, a Administração Pública, a sociedade; do outro, o Mercado, a iniciativa particular e dos indivíduos o que torna essencial a exposição desses conceitos.

Para uma melhor compreensão da pesquisa e de sua área de aplicação, faz-se importante apresentar algumas definições sobre esse tipo de organização: qual o seu regramento jurídico, seu enquadramento societário e tributário e sua relação com as demais organizações públicas e privadas.

O terceiro setor pode ser compreendido com um campo amplo, diverso e por vezes contraditório, composto por organizações com diferentes perfis culturais, ideológicos e operacionais.

Conforme Pereira (2000), trata-se de um setor heterogêneo, formado por entidades que vão desde iniciativas progressistas voltadas à promoção da justiça social e emancipação popular, até organizações com caráter assistencialista e estratégias alinhadas à lógica do mercado.

Apesar das diferenças, um elemento comum é a constante referência ao discurso da cidadania. O novo associativismo característico do terceiro setor estabelece relações ambíguas com o associativismo tradicional, oriundo dos movimentos sociais urbanos das décadas de 1970 e 1980.

Segundo Fernandes, citado por Paes (2019),

É necessário esclarecer que, junto com Estado (Primeiro Setor) e com o Mercado (Segundo Setor), identifica-se a existência de um Terceiro Setor, mobilizador de um grande volume de recursos humanos e pessoais para impulsionar iniciativas voltadas para o desenvolvimento social, setor no qual se inserem as sociedades civis sem fins lucrativos, as associações e as

fundações de direito privado, todas entidades de interesse social (Paes, 2019, p. 76).

Nestas organizações do Terceiro Setor, não há a distribuição de resultados financeiros, mas há a relação de agência entre os gestores da organização (agentes) e os seus doadores e voluntários (principais).

Essa relação agente-principal presente em qualquer organização “é sempre conflituosa quando um determinado individuo (agente) age em nome de outro (principal) e os objetivos de ambos não coincidem integralmente” (Mendonça; Machado, 2004, p. 303).

Nota-se então que as entidades do terceiro setor podem ser tratadas como uma união de esforços entre pessoas físicas e jurídicas, públicas ou privadas com o intuito de criar entidades para atender as necessidades da sociedade, em áreas, temas e ações em que o poder público não consegue alcançar.

O Código Civil Brasileiro em seu art. 40 da Lei nº 10.406/2002 (BRASIL, 2002), em seu título II que trata das “pessoas jurídicas”, define que “as pessoas jurídicas são de direito público, interno ou externo, e de direito privado”.

O art. 41 da Lei nº 10.406/2002 (BRASIL, 2002), define que “são pessoas jurídicas de direito público interno: a União, os Estados, o Distrito Federal e os Territórios, os Municípios; as autarquias, inclusive as associações públicas e as demais entidades de caráter público criadas por lei”.

No art. 44 da Lei nº 10.406/2002 (BRASIL, 2002), encontramos a definição “das pessoas jurídicas de direito privado: as associações, as sociedades, as fundações, as organizações religiosas, e os partidos políticos”.

No art. 62 da Lei 10.406/2002, determina que:

Para criar uma fundação, o seu instituidor fará, por escritura pública ou testamento, dotação especial de bens livres, especificando o fim a que se destina, e declarando, se quiser, a maneira de administrá-la. Parágrafo único: A fundação somente poderá constituir-se para fins de: assistência social, cultura, defesa e conservação do patrimônio histórico e artístico, educação, saúde, segurança alimentar e nutricional, defesa, preservação e conservação do meio ambiente e promoção de desenvolvimento sustentável, pesquisa científica, desenvolvimento de tecnologias alternativas, modernização de sistemas de gestão, produção e divulgação de informações e conhecimentos técnicos e científicos, promoção da ética, da cidadania, da democracia e dos direitos humanos, atividades religiosas (BRASIL, 2002, p. 7).

No art. 66 da Lei nº 10.406/2002 (BRASIL, 2002), “velará pelas fundações o Ministério Público do Estado onde situadas”. Como determina o regramento legal, a criação de uma fundação é um ato de solidariedade de uma pessoa física ou jurídica, que por sua decisão resolve dispor de alguns ou da totalidade de seus bens, para fins sociais específicos.

Além de tudo pode-se destacar um ato de amor para com toda a sociedade. Esses bens deverão ser utilizados no objetivo específico destinado no estatuto da fundação.

O Estado também pode realizar a criação de fundação pública de direito privado – por meio de lei – com o intuito de atender a demandas e necessidades da sociedade em geral.

De acordo com Di Pietro (1998), quando o estado:

Institui pessoa jurídica sob forma de fundação, ele pode atribuir a ela regime jurídico administrativo, com todas as prerrogativas e sujeições que lhe são próprias, ou subordiná-la ao Código Civil; neste último caso, com derrogações por normas de direito público” (Di Pietro 1998, p. 322-323).

Segundo Paes (2017), pode-se classificar as fundações nos seguintes tipos:

1. Fundações instituídas e/ou mantidas pelo Poder Público:
 - a. Com personalidade jurídica de direito privado;
 - b. Com personalidade jurídica de direito público (autarquias);
2. Fundações de direito privado:
 - a. Instituídas por pessoas físicas e/ou jurídicas;
 - b. Instituídas por empresas;
 - c. Instituídas por partido político;
 - d. De apoio a instituições de ensino superior;
 - e. De previdência privada ou complementar;
 - f. Comunitárias (Paes, 2017, p. 180).

As fundações podem ser classificadas de diversas formas, quanto ao modo como foram instituídas e de acordo com o seu objetivo específico de atuação e se enquadram como fundações de direito privado, criadas com o objetivo de apoio a instituições de ensino superior, que são o objetivo desta pesquisa. Será apresentado no próximo capítulo, a definição e o regramento legal dessas organizações.

2.1.1 Fundação de Apoio

Neste tópico, serão discutidos os principais conceitos relacionados às fundações de apoio, abordando o que são, sua finalidade, os objetivos de sua constituição, entre outras informações relevantes.

A fundação de apoio, trata-se de um ente de direito privado, sem fins lucrativos e que por característica deste tipo societário é criada em sua maioria por pessoas físicas (geralmente professores), pessoas jurídicas (instituições ou organizações de educação de ensino superior) e a própria universidade (Paes, 2017, p. 216).

Tem por objetivo fomentar os projetos de ensino, pesquisa e extensão das universidades federais e demais instituições de ensino superior públicas ou privadas e as Instituições Científicas e Tecnológicas. (Paes, 2017, p. 216)

Para tratar efetivamente deste tema, foi publicada pela União a lei nº 8.958/1994 (BRASIL, 1994), que “Dispõe sobre as relações entre as instituições federais de ensino superior e de pesquisa científica e tecnológica e as fundações de apoio e dá outras providências” e trata especificamente dessas organizações e das atividades que podem ser realizadas por elas.

A lei nº 8.958/1994 (BRASIL, 1994), em seu art. 1º enfatiza:

As Instituições Federais de Ensino Superior – IFES e as demais Instituições Científicas e Tecnológicas – ICTs, poderão celebrar convênios e contratos, ..., por prazo determinado, com as fundações instituídas com a finalidade de apoiar aos projetos de ensino, pesquisa, extensão, desenvolvimento institucional, científico e tecnológico e estímulo a inovação, inclusive na gestão administrativa e financeira necessária à execução destes projetos (BRASIL, 1994, p. 1).

Esse primeiro artigo da Lei apresenta o que as fundações de apoio fazem no auxílio ao desenvolvimento da educação e do ensino superior brasileiro além disso demonstra que será responsável por toda a gestão administrativa e financeira destes projetos.

Exceto, ao que tange os critérios técnicos que ficam a cargo da instituição apoiada através do docente que será indicado como coordenador do projeto junto a fundação de apoio.

A fundação como qualquer outra organização, deverá manter os fluxos, processos e atividades mapeadas e em conformidade com o que determina a legislação societária e tributária, além de auxiliar o coordenador na execução desses convênios e contratos com fim de atender ao objetivo desses projetos.

O art. 3º-A, desta mesma lei determina que:

Na execução de convênios, contratos, acordos e demais ajustes na forma desta Lei, as fundações de apoio deverão: prestar contas dos recursos aplicados aos entes financiadores; submeter-se ao controle de gestão pelo órgão máximo da IFES e submeter-se ao controle finalístico pelo órgão de controle governamental competente (BRASIL, 1994, p. 3).

Esse artigo evidencia que a fundação de apoio deve possuir seus controles internos e externos eficientes, em toda a execução administrativa e financeira do projeto garantindo à fundação, ao coordenador do projeto e à instituição apoiada que os recursos serão aplicados e geridos dentro do que determina o plano de trabalho e as regras legais dos órgãos fiscalizadores.

Em continuidade esse artigo observa que as fundações apoio são fiscalizadas e controladas em todas as ações e atividades realizadas enquanto gestoras de projetos, seja pela instituição apoiada, seja pelo financiador do projeto, seja pelo seu Conselho Fiscal e Diretor ou pelo próprio Ministério Público de contas, na análise de sua prestação de contas anual.

Ao analisar a missão, visão e valores disponíveis no portal eletrônico da Universidade Federal do Paraná (UFPR), é possível destacar a importância do ensino superior público em nosso país.

A missão da UFPR é “produzir, disseminar, fomentar e aplicar o conhecimento, contribuindo para a construção de uma sociedade inclusiva, equânime, solidária e para o desenvolvimento sustentável”.

A visão busca “consolidar e ampliar a condição de Universidade de expressão internacional em ensino, pesquisa, extensão e inovação, abrigando iniciativas científicas e culturais voltadas para a promoção da cidadania e da soberania nacional”.

Por fim os seus valores seguem a “excelência em ensino, pesquisa e extensão, valorização da ciência, ética pública e institucional, criatividade e inovação, desenvolvimento institucional e social, cidadania e inclusão, sustentabilidade, projeção e integração internacional”.

As fundações de apoio às IFES e ICTs foram criadas com o intuito de auxiliar essas instituições a cumprirem com sua missão institucional, e aparecem como um braço forte de apoio nas demandas da gestão administrativa e financeira dos projetos, garantindo assim a captação de recursos, para o investimento em ensino, pesquisa, extensão, desenvolvimento institucional e tecnológico.

Conforme relatório disponível no portal eletrônico do MEC, existem 106 fundações de apoio credenciadas ou autorizadas para prestar apoio de gestão de projetos às IFES.

Isso demonstra que as fundações são um importante elo entre a sociedade e a comunidade universitária, buscando a captação de recursos disponíveis junto a agências de fomento, e ao próprio setor público e privado para desenvolvimento cada vez maior do ensino, pesquisa, extensão, desenvolvimento institucional e inovação.

2.2 ANÁLISE EXPLORATÓRIA DE DADOS (AED)

Neste tópico o objetivo é apresentar as principais descobertas do levantamento bibliográfico, sobre a análise exploratória de dados e apresentar suas principais técnicas que serão aplicadas nessa pesquisa.

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), a análise exploratória de dados desempenha um papel essencial na descoberta de conhecimento, auxiliando gestores na extração de informações estratégicas para a otimização de processos.

A AED busca examinar os dados coletados e compreender as relações entre as variáveis antes de sua aplicação em métodos estatísticos (Ferreira *et al.*, 2021). Segundo Sicsú, Samartini e Barth (2023), a AED permite identificar padrões ocultos, *outliers* e correlações, garantindo uma base sólida para a construção de modelos e a proposição de melhorias.

A gestão de dados tem se tornado um fator estratégico nas organizações e são base para a tomada de decisão. Nesse contexto, o trabalho pretende discutir a aplicação de técnicas de AED em conjunto com a mineração de dados.

Com o objetivo de encontrar tendências e padrões que sejam úteis para auxiliar o gestor da fundação de apoio na tomada de decisões estratégicas na captação de recursos melhorando a sustentabilidade financeira da instituição.

Segundo (Ferreira *et al.*, 2021), existem diversas estratégias utilizadas para diferentes abordagens estatísticas, são elas conforme destacam o Quadro 1:

QUADRO 1 – ABORDAGENS ESTATÍSTICAS E ESTRATÉGIAS UTILIZADAS

Abordagem	Estratégia
Estatística clássica	Problema – Dados – Modelo – Análise
Estatística bayesiana	Problema – Dados – Modelo Priori – Análise
Análise exploratória de dados	Problema – Dados – Análise – Modelo

Fonte: Adaptado de Ferreira *et al.* (2024).

Diferentemente da abordagem clássica e bayesiana, que prioriza a criação do modelo antes da análise, a AED foca na exploração e mineração dos dados para identificar padrões e percepções, definindo posteriormente o modelo mais adequado (Ferreira *et al.*, 2021). Esse método permite uma compreensão mais aprofundada das informações sem uma redução significativa da quantidade de dados (Ferreira *et al.*, 2021).

Para Stankevix (2020), a análise exploratória de dados representa uma etapa central no processo analítico, pois organiza, resume e aplica técnicas estatísticas de forma a tornar as informações mais compreensíveis visualmente.

O trabalho da AED está diretamente ligado a estatística descritiva, sendo dividido em etapas: coleta, organização, tratamento, análise, apresentação e interpretação dos dados, pois as etapas de organização, tratamento, análise e apresentação de dados utilizam técnicas descritivas (Ferreira *et al.*, 2021).

Segundo as interpretações dos autores a AED pode ser uma ferramenta muito importante para as organizações para que os gestores consigam ter informações de forma útil.

Cada vez mais rápida e consolidada, trazendo percepções e caminhos a seguir que em muitas vezes estão ocultos dentro dos dados, o que não seria possível sem a aplicação dessas técnicas.

2.3 MINERAÇÃO DE DADOS

Neste tópico, o objetivo é apresentar os principais achados do levantamento bibliográfico, com ênfase nas abordagens dos autores sobre a mineração de dados e o propósito por trás da aplicação dessa poderosa técnica.

Sobre a definição de mineração de dados, são diversos os pontos em comum entre os autores citados a seguir, mas o mais importante refere-se às possibilidades

de melhoria para as organizações na tomada de decisão estratégica que são baseadas em mineração de dados.

Segundo Castro e Ferrari (2016, p. 25), “o processo de mineração corresponde à extração de minerais valiosos, como ouro e pedras preciosas de uma mina”.

Ao fazer alusão a minerais valiosos e preciosos, o autor nos lembra que, embora esses minerais não possam ser cultivados ou produzidos artificialmente, eles podem ser extraídos, desde que haja acesso à mina e ferramentas apropriadas para a mineração e extração, seguidas da comercialização (Castro; Ferrari, 2016).

É com base nessa analogia que surge o termo "mineração de dados", que envolve a exploração de uma base de dados (mina) utilizando algoritmos (ferramentas) específicos para obter conhecimento (minerais preciosos) (Castro; Ferrari, 2016).

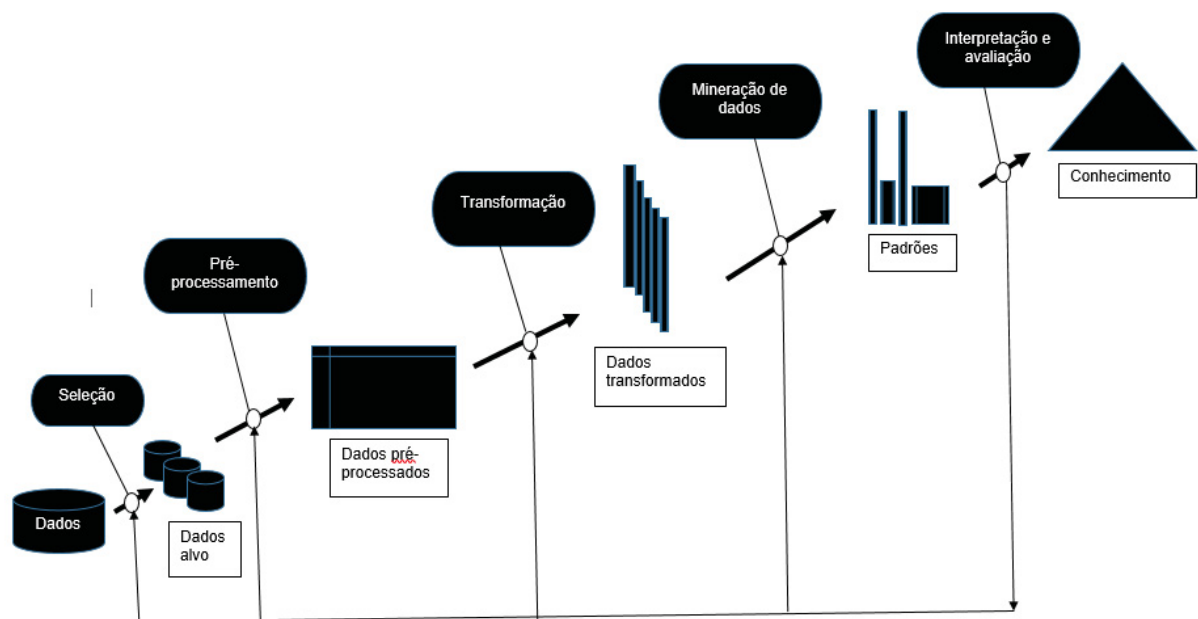
Enquanto os dados são símbolos ou signos não estruturados, sem significado por si só, como valores em uma tabela, a informação emerge das descrições, agregando sentido e utilidade aos dados (Castro; Ferrari, 2016).

Nesse sentido, autores afirmam que o conceito de mineração de dados está inserido no contexto do processo de descoberta de conhecimento em bases de dados (KDD), que é o termo mais abrangente.

A mineração de dados é definida como a etapa específica de descoberta dentro do ciclo KDD, que envolve desde o pré-processamento dos dados (seleção, limpeza e transformação) até a avaliação final dos padrões encontrados (Castro; Ferrari, 2016).

Segundo Fayyad, Piatetsky-Shapiro e Smith (1996), KDD é um processo não-trivial, de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis, implícitos nos dados, como pode ser observado na Figura 1, as fases do processo de KDD.

FIGURA 1 – FASES DO PROCESSO KDD.



Fonte: Adaptada de Fayyad, Piatetsky-Shapiro e Smith (2024).

O processo de descoberta do conhecimento em bases de dados (KDD) é um ciclo multifásico que se inicia com a compreensão do domínio e dos dados, passando pela seleção, limpeza e transformação e, finalmente, a aplicação da mineração de dados. O ciclo se completa com a interpretação, avaliação dos padrões extraídos e a aplicação do conhecimento na tomada de decisão (Fayyad; Piatetsky-Shapiro e Smith, 2024).

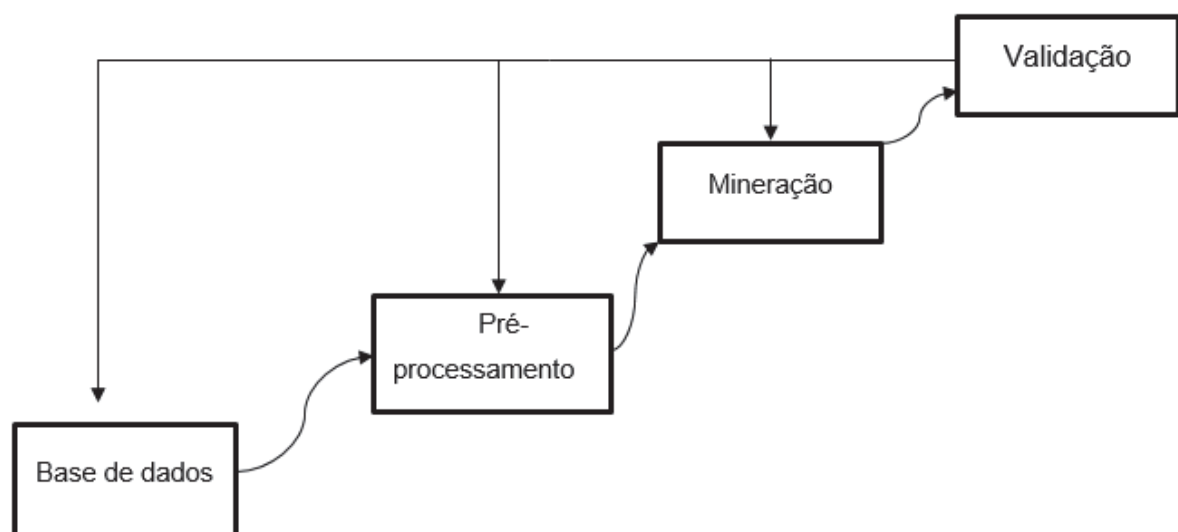
Embora o processo de descoberta de conhecimento em bases de dados (KDD) seja detalhado em diversas subetapas, ele pode ser sintetizado em quatro fases principais: a **Base de dados** (coleta e organização dos dados brutos), a **Preparação ou Pré-processamento** (limpeza, integração e transformação dos dados), a **Mineração de Dados** (aplicação dos algoritmos de extração do conhecimento, como agrupamento e classificação) e, por fim, a **Avaliação ou Validação do Conhecimento** para identificar padrões úteis e não triviais (Castro; Ferrari, 2016, p.26-27).

As quatro etapas do processo de mineração de dados são correlacionadas e interdependentes, o que torna essencial considerar as inter-relações entre elas e sua influência no resultado para a extração eficaz de informações relevantes.

Esse processo deve permitir a extração e validação de conhecimentos úteis sob diferentes perspectivas, os quais podem ser aplicados para a tomada de decisões estratégicas em diversas áreas, como controle de processos, gestão da informação e conhecimento, processamento de consultas, entre outras aplicações (Castro, Ferrarini, 2016).

Abaixo é possível observar na Figura 2, os passos da KDD, segundo Castro e Ferrarini (2016).

FIGURA 2 – PASSOS DA KDD



Fonte: Adaptado de Castro, Ferrarini (2016).

Tanto Fayyad, Piatetsky-Shapiro e Smith (1996), quanto Castro e Ferrari (2016), destacam a importância do processo de KDD como uma sequência de etapas interdependentes e não-triviais, visando a extração de padrões válidos e úteis a partir de grandes volumes de dados. O processo envolve desde a compreensão do problema e a seleção adequada dos dados até a aplicação do conhecimento adquirido para ações práticas.

A eficácia desse processo depende de uma análise cuidadosa em cada fase, como a preparação dos dados, a escolha de técnicas de mineração, a validação dos resultados e a adaptação desses conhecimentos ao contexto específico do cliente ou da organização.

Assim, o KDD é essencial para transformar dados brutos em informações estratégicas, que podem contribuir para decisões mais informadas e eficientes em diversas áreas.

Segundo Goldschmidt (2015), a etapa de pré-processamento envolve diversas funções essenciais, como a captação, organização, tratamento e preparação dos dados para a mineração.

Essa fase é de grande importância no processo de descoberta de conhecimento, pois abrange a correção de erros e inconsistências nos dados, além do ajuste da formatação necessária para que os algoritmos de mineração possam ser aplicados de forma eficaz. Esta etapa tem por objetivo a preparação dos dados para os algoritmos que serão aplicados na etapa de Mineração de Dados.

A etapa de pré-processamento e transformação dos dados será utilizada nessa pesquisa nos arquivos gerados através do banco de dados do sistema de gestão da fundação, seguindo os passos conforme descritos pelos autores consultados.

Segundo Goldschmidt (2015), a etapa de Mineração de Dados é a principal fase do processo KDD, sendo dedicada à busca por conhecimento úteis e implícitos nos dados.

A mineração de dados envolve tarefas como associação, classificação, regressão, agrupamento, sumarização, detecção de desvios e descoberta de sequências, aplicadas para identificar padrões, prever valores, segmentar registros e descrever características, utilizando algoritmos e técnicas específicas (Fayyad et al., 1996; Goldschmidt, 2015).

Goldschmidt (2015) aponta que várias tarefas de mineração de dados podem ser adaptadas ou combinadas para originar novas tarefas, mais complexas e úteis no contexto da KDD. Um exemplo é a adaptação da tarefa de Descoberta de Associações para incluir a Mineração de Regras de Associação Generalizadas (Srikant et al., 1997).

Além disso, a composição de tarefas como agrupamento (clusterização) e classificação é frequente, permitindo que os *clusters* identificados sejam posteriormente mapeados por uma função de classificação, potencializando a análise (Han e Kamber, 2012; Weiss e Indurkha, 1998).

Ainda segundo o autor, a qualidade dos dados desempenha um papel crucial na eficácia dos modelos de conhecimento gerados a partir do processo de KDD. Dados de baixa qualidade inevitavelmente resultam em modelos de conhecimento igualmente insatisfatórios, um princípio frequentemente resumido pelo acrônimo

GIGO (do inglês, *Garbage in Garbage out*), que destaca a dependência direta entre a qualidade da entrada e o resultado obtido.

A base de dados tem uma importância fundamental no processo de descoberta de conhecimento em banco de dados e mineração de dados e exige um cuidado rigoroso com a qualidade dos dados. A escolha adequada das técnicas de análise e a compreensão das tarefas específicas para extrair o conhecimento desejado.

2.4 TOMADA DE DECISÃO

Neste tópico, serão abordadas as considerações e achados dos autores sobre a tomada de decisões e sua importância para a estratégia das organizações e empresas dos diversos níveis decisórios.

De acordo com Marques e Correia Neto (2020), a tomada de decisão pode variar em termos de complexidade, duração, compartilhamento e formalidade, sendo um processo fundamental que é tão relevante quanto a decisão propriamente dita.

Conforme Yu (2011), a tomada de decisão é um dos momentos mais críticos na administração de qualquer empreendimento humano, sendo influenciada por limitações humanas, aspectos sociais como afetividade e relacionamentos, além da complexidade organizacional.

Algumas pessoas se preparam de forma detalhada para esse processo, enquanto outras optam por confiar na intuição ou esperar por uma solução espontânea para o problema ou dúvida em questão.

Ainda de acordo com Marques e Correia Neto (2020) no processo de tomada de decisão, é crucial que a decisão seja bem fundamentada, especialmente em situações em que pode acarretar riscos significativos à organização. Uma tomada de decisão baseada em informações facilita o gestor a encontrar a melhor solução ou pelo menos mitigar os riscos e prejuízo para a organização.

Tomar decisões é uma atividade constante em diversas esferas, incluindo o ambiente organizacional, onde o foco está em buscar soluções que agreguem valor às partes interessadas.

A qualidade das decisões está diretamente relacionada à eficácia dos processos que conduzem a esses resultados, sendo importante garantir rapidez e

precisão nos julgamentos para alcançar decisões de alta qualidade (Abu-Alsodos, 2023).

Ainda segundo Marques e Correia Neto (2020), as decisões podem ser classificadas de diversas maneiras, como individuais ou coletivas, urgentes ou que permitem tempo para análise, e com base no custo para alcançar a solução.

Além disso, existem fatores que influenciam diretamente o processo decisório, como o tempo disponível, os recursos à disposição, o perfil do decisor, os aspectos comportamentais e o contexto organizacional.

As decisões podem ser afetadas diretamente por barreiras de natureza crítica das decisões, pois as decisões podem representar um impacto alto, são em sua maioria complexas, geram incerteza e podem ser vinculadas a questões temporais com prazo para a tomada de decisão.

Além disso, entram nesse quesito as barreiras cognitivas e humanas que podem afetar diretamente a forma como tomamos as decisões, que podem depender da forma que pensamos, de nossa cultura, crenças e aprendizados e muitas vezes podem levar a decisões de forma rápida e precipitada.

O processo de tomada de decisão pode seguir alguns passos como identificar e compreender o problema, listar e definir os objetivos, obter as informações, considerar questões morais e éticas, gerar e analisar as alternativas, utilizar métodos analíticos e, criticar os resultados obtidos (Marques; Correia Neto, 2020).

Os autores demonstram a importância da tomada de decisões em diversos contextos, destacando que ela deve ser bem fundamentada, considerando tanto os fatores internos quanto externos, e seguir um processo estruturado para minimizar riscos e maximizar os resultados.

Segundo Rezende e Abreu (2013, p. 76), “as informações devem ser úteis às decisões, possuem conteúdo único, exigem mais de duas palavras...”. Para Davenport e Prusak (2003, p. 4), a informação tem por finalidade “mudar o modo de como o destinatário vê algo, exercer algum impacto sobre o seu julgamento e comportamento. A informação não só ‘dá forma’ ao receptor como visa alguma finalidade”.

Choo (2003, p. 30), “durante a tomada de decisões, a principal atividade é o processamento e a análise da informação a partir das alternativas disponíveis, cujas vantagens e desvantagens são pesadas”.

Ainda segundo Choo (2003), a organização:

Depois que criou significados e construiu conhecimento para agir, precisa escolher entre várias opções ou capacidades disponíveis e se comprometer com uma única estratégia. E complementa que todos os comportamentos da organização nascem de decisões, as características essenciais da estrutura organizacional derivam de característica do processo decisório e da escolha racional humana (Choo, 2003, p. 41).

Para Stoner e Freeman (1999, p. 182), a tomada de decisão é “o processo de identificar um problema específico e selecionar uma linha de ação para resolvê-lo”.

A tomada de decisão organizacional é formalmente definida como o processo de identificação e solução de problemas. O processo contém dois estágios principais. No estágio de identificação do problema, a informação sobre as condições ambientais e organizacionais é monitorada para determinar se o desempenho é satisfatório e para diagnosticar a causa das falhas. O estágio de solução do problema se dá quando os caminhos alternativos de ação são considerados e uma alternativa é selecionada e implementada. (Daft, 2006, p. 372).

Ainda segundo Huang e Yue (2024), o primeiro passo no processo de tomada de decisão é identificar problemas e definir metas estratégicas que podem envolver a análise de mercado, uma avaliação de ambiente competitivo e uma revisão dos recursos e capacidades da empresa.

Já para Moussas (2024), os líderes podem descobrir *insights* profundos, permitindo decisões informadas que podem refinar processos e fortalecer a posição competitiva de uma organização no mercado.

De acordo com Abu-Alsondos (2023), a capacidade de tomar decisões estratégicas com sucesso permite que a empresa mantenha uma vantagem competitiva e alinhe seus processos internos com seu ambiente externo. Além disso, permite que as empresas suportem dificuldades e riscos. Por outro lado, más escolhas estratégicas resultam no colapso de uma organização.

A capacidade de tomar decisões estratégicas eficazes permite que as empresas mantenham vantagem competitiva e superem desafios, enquanto decisões mal formuladas podem levar ao fracasso organizacional (Abu-AlSondos, 2023).

Com base nos resultados da análise de dados, as empresas podem fazer recomendações de otimização e recomendações estratégicas para melhorar a sua tomada de decisão e operações comerciais (Aghuir; Dusa; Onatã, 2019).

Segundo Huang e Yue (2024), com base na coleta e análise de dados, as empresas geram várias opções estratégicas e podem usar diferentes métodos de avaliação para determinar a fase de tomada de decisão que envolve o

desenvolvimento formal e o planejamento da decisão, incluindo a alocação de recursos e de responsabilidades. O estágio de implementação da decisão é o processo de colocar toda a estratégia em prática.

A partir do momento que a empresa tem bem claro os problemas e seus objetivos, a empresa coleta uma variedade de dados, incluindo dados de mercado, dados financeiros, de vendas, de compras, dados de clientes etc., que serão usados para analisar e avaliar diferentes opções estratégicas (Huang; Yue, 2024).

No contexto atual, segundo Huang e Yue (2024), a análise de dados desempenha um papel crucial na tomada de decisões estratégicas. A coleta de dados e a avaliação de opções são fundamentais para definir metas e implementar estratégias.

Dessa forma, o sucesso da organização está intimamente ligado à sua capacidade de tomar decisões rápidas, precisas e bem fundamentadas, garantindo alinhamento com o ambiente competitivo e otimizando suas operações.

Os autores destacam a importância da análise dos dados e da qualidade das informações no processo decisório, especialmente na escolha das estratégias que visam alcançar o sucesso organizacional.

A tomada de decisão, no contexto organizacional, é um processo complexo e estruturado que depende essencialmente da coleta, análise e avaliação criteriosa dos dados. A capacidade de interpretar e utilizar as informações de maneira eficaz é crucial para fundamentar a escolha das alternativas estratégicas.

Essa análise deve considerar tanto o ambiente interno da organização quanto os fatores externos, assegurando que as decisões tomadas estejam alinhadas com os objetivos da empresa e com as demandas do mercado, o que é fundamental para garantir o sucesso a longo prazo.

Goldshmidt (2015) discute que o valor dos dados armazenados está diretamente relacionado à capacidade de extrair conhecimento mais avançado a partir deles, ou seja, transformar os dados em informações úteis para apoiar a tomada de decisões e/ou para compreender melhor o fenômeno que originou os dados.

O autor também destaca a importância de entender a hierarquia entre dados, informações e conhecimento. Nesse contexto, os dados são descritos como cadeias de símbolos sem significado, enquanto a informação é composta por dados processados, com significados e contextos definidos.

O conhecimento, por sua vez, é visto como um padrão ou conjunto de padrões que podem envolver e conectar dados e informações, gerando percepções valiosas para a organização. Dessa forma, informação e conhecimento formam a base para a tomada de decisões em diversos cenários.

Os dados processados transformam-se em informações, que serão utilizados pelos gestores na tomada de decisão estratégica da fundação, para definir qual planejamento será seguido para a estruturação de ações e metas de captação de recursos baseadas em dados.

2.5 GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO

Neste tópico, serão abordadas as considerações e achados dos autores sobre a gestão da informação e do conhecimento e sua importância para a estratégia das organizações e empresas e sua relação com os diversos níveis decisórios

A GI se concentra na estrutura para coletar e armazenar os dados, enquanto a GC foca na interpretação, disseminação e internalização do conhecimento (Takeuchi; Nonaka, 2008; Tarapanoff, 2006).

O sucesso de uma organização na era da informação depende de sua capacidade de gerenciar o fluxo de dados e transformá-los em informações úteis. A Gestão da Informação (GI) e a Gestão do Conhecimento (GC) são processos fundamentais que garantem essa transição.

O processo que guia a transformação de insumo bruto em subsídio gerencial é definido pela Hierarquia DIKW (Dados, Informação, Conhecimento e Sabedoria), que estabelece a progressão de valor desde os Dados até a Decisão. Esta progressão é sumarizada no Quadro 2, que mapeia a relação entre os conceitos e a funcionalidade prática no contexto da pesquisa.

QUADRO 2 – FLUXO DE GERAÇÃO DE VALOR: DA BASE DE DADOS AO SUPORTE À DECISÃO

Nível de Abstração	Componente (DIKW)	Conceituação e Processo	Autores citados na Dissertação
1. Insumo Bruto	Dados	São cadeias de símbolos sem significado inerente, o nível mais básico de abstração, sendo a coleção	Goldshmidt (2015), Castro e Ferrari (2016).

		organizada que permite a recuperação eficiente.	
2. Processamento	Informação	Dados que foram processados, estruturados e contextualizados, ganhando significado. Tem a finalidade de exercer impacto sobre o julgamento do destinatário.	Goldshmidt (2015), Davenport e Prusak (2003).
3. Valor/Padrão	Conhecimento	Padrão ou conjunto de padrões que conectam dados e informações, sendo extraído a partir dos dados pré-processados por meio da aplicação de algoritmos de mineração.	Goldshmidt (2015), Castro e Ferrari (2016).
4. Resultado Gerencial	Decisão	É o processo de identificar um problema e selecionar uma linha de ação para resolvê-lo. É o momento mais crítico da administração, sendo principal atividade o processamento e análise da informação/conhecimento.	Stoner e Freeman (1999), Choo (2003), Marques e Correia Neto (2020).

Fonte: O Autor (2025).

De acordo com Razzolini Filho (2009), a gestão da informação é conceituada como um processo que envolve diversas atividades, tais como a busca, identificação, processamento, armazenamento e disseminação de informações.

Segundo Teixeira e Freitas (2014), a informação é conceituada como uma mensagem com o objetivo de provocar alguma mudança no receptor ou alterar a sua perspectiva.

No contexto da gestão da informação (GI), ela está estruturada de forma a garantir a recuperação eficaz e em tempo hábil das informações, o que auxilia diretamente no processo de tomada de decisão. No ambiente empresarial, muitas organizações não têm pleno conhecimento das informações que já possuem ou das que necessitam.

A GI, em seu processo de busca, envolve a seleção de fontes confiáveis e de qualidade para atender ao usuário, e segue critérios como: classificação, que facilita

o tratamento, processamento e recuperação de informação; processamento, que visa tratar a informação e tornar sua compreensão mais acessível; armazenamento, que utiliza técnicas de arquivamento para garantir o acesso quando necessário; disseminação, que assegura que a informação transite no momento adequado, criando utilidade temporal; e recuperação, que consiste na definição de mecanismos para permitir a localização da informação armazenada.

Não é suficiente apenas acesso à informação, é fundamental saber como buscá-la e o que exatamente buscar. Além disso, após a identificação da informação, é essencial processá-la corretamente e compreender como ela pode responder a questionamentos ou resolver problemas.

Outro aspecto crucial é o armazenamento e a disseminação dessas informações, garantindo que possam ser reutilizadas por toda a organização. Isso contribui para a criação de uma cultura de compartilhamento de informação e conhecimento, promovendo um ambiente colaborativo e eficiente.

De acordo com Teixeira e Freitas (2014), o uso estratégico da informação tem ganhado destaque nas organizações, uma vez que as decisões estratégicas envolvem o planejamento de longo prazo, a busca por vantagem competitiva, a adaptação a mudanças no ambiente organizacional e o estabelecimento de recursos e competências necessárias para alcançar os objetivos e atender às expectativas dos acionistas.

A informação, por sua natureza contribui para o desenvolvimento estratégico organizacional, pois proporciona aos gestores uma visão ampla dos processos e das inter-relações entre os departamentos.

Nesse contexto, o processo de administração estratégica envolve um conjunto de compromissos, decisões e ações essenciais para garantir que a empresa alcance vantagem competitiva

Teixeira e Freitas (2014), McGee e Prusak (1994) e Beuren (2000) identificaram as etapas do gerenciamento estratégico da informação que são: identificação da necessidade de informação, classificação e armazenamento da informação, distribuição e disseminação da informação e análise e uso da informação.

Ainda conforme Teixeira e Freitas (2014), a análise das etapas para a gestão estratégica da informação revela que a informação está presente em todos os processos de gestão, desempenhando um papel crucial para a sobrevivência e o sucesso das organizações.

A informação estratégica é o resultado da interpretação dos dados externos ao ambiente e dos aspectos internos da organização, fornecendo ao gestor as bases necessárias para tomar decisões e definir os rumos que permitem cumprir a missão organizacional (Teixeira; Freitas, 2014).

A informação estratégica não é apenas um recurso valioso, mas um pilar essencial para o desenvolvimento e a sustentabilidade das organizações no cenário atual.

Segundo Takeuchi e Nonaka (2008), a humanidade passou, em menos de duzentos anos, de uma sociedade industrial para a sociedade do conhecimento, o que trouxe uma mudança no entendimento sobre paradoxos. Se antes eram vistos como elementos a serem evitados, passaram a ser reconhecidos como algo a ser aceito e cultivado.

Os autores afirmam que as contradições, dicotomias e dualidades fazem parte da própria essência do conhecimento, que se compõe de dois elementos aparentemente opostos: o conhecimento explícito e o tácito (Takeuchi; Nonaka, 2008).

O primeiro pode ser facilmente comunicado e formalizado, enquanto o segundo é mais pessoal e difícil de compartilhar, pois está profundamente enraizado nas experiências individuais e em aspectos como valores e emoções. Ainda segundo os autores, o conhecimento tácito possui duas dimensões principais: uma técnica, que envolve habilidades informais e muitas vezes associadas ao conceito *know-how* e outra cognitiva, que abrange crenças, valores, percepções, e modelos mentais (Takeuchi; Nonaka, 2008).

De acordo Takeuchi e Nonaka (2008), as organizações criam e utilizam conhecimento por meio de um processo contínuo de conversão entre as formas do conhecimento tácito e explícito. Os autores identificam quatro modos de conversão: socialização (de tácito para tácito), externalização (de tácito para o explícito), combinação (de explícito para explícito) e internalização (de explícito para tácito).

Esse ciclo é conhecido como o modelo SECI, que descreve como o conhecimento, tanto tácito quanto explícito, é ampliado em termos de qualidade e quantidade, e transita do indivíduo para o grupo e, posteriormente, para a organização (Takeuchi; Nonaka, 2008).

O modelo SECI envolve o processo de espiral, no qual cada modo de conversão do conhecimento ocorre em diferentes níveis: socialização (de indivíduo

para indivíduo), externalização (de indivíduo para grupo), combinação (de grupo para indivíduo) e internalização (de organização para indivíduo) (Takeuchi; Nonaka, 2008).

De acordo com Drummond (2008), a Gestão do Conhecimento (GC) envolve um conjunto de atividades focadas na promoção do conhecimento organizacional, permitindo que as organizações e seus colaboradores utilizem as informações e conhecimentos mais adequados para alcançar seus objetivos e aumentar a competitividade no mercado.

Ainda conforme Drummond (2008), Davenport e Prusak (1998) destacam que a principal vantagem competitiva sustentável de uma empresa está relacionada ao conhecimento coletivo à eficiência no uso desse conhecimento e à capacidade de adquirir e praticar novos saberes com agilidade.

De acordo com Choo (2003), citado por Drummond (2008), as organizações do conhecimento utilizam estrategicamente a informação e o conhecimento para atuar em três arenas distintas, porém imbricadas: construção de sentido, criação do conhecimento e tomada de decisão.

A eficácia no processo de criação do conhecimento está atrelada a um contexto capacitante, entendido como um espaço compartilhado que estimula o surgimento de novos relacionamentos.

Esse conceito, denominado “ba” (ou lugar/espço do conhecimento), pode se manifestar de forma física, virtual, mental ou, na maioria das vezes, por meio da combinação desses três aspectos (Von Krogh Ichijo; Nonaka, 2001, apud Drummond, 2008)

Ainda segundo Drummond (2008), a GC assume quatro diferentes perspectivas: estratégicas, tático-gerencial, perspectiva tecnológica e perspectiva de gestão da mudança e a tríade da GC está baseada na gestão da informação, tecnologia da informação e conversões entre conhecimentos organizacionais tácitos e explícitos.

Criar um ambiente propício para o compartilhamento e a disseminação de informações e conhecimento pode oferecer diversas vantagens competitivas para qualquer organização.

Esse processo transforma a informação em um ativo estratégico, possibilitando a geração de conhecimento em toda a estrutura organizacional, otimizando a tomada de decisões e contribuindo para o desenvolvimento da inteligência organizacional.

Dessa forma, apoia-se a definição de estratégias, metas e direcionamentos para o sucesso nos negócios. Este tema está diretamente relacionado ao projeto de pesquisa, que visa identificar e analisar informações na base de dados de uma fundação de apoio, transformando esses dados em informações estratégicas e úteis.

O objetivo é descobrir conhecimentos relevantes que possam ser convertidos em vantagem estratégica, promovendo a geração de conhecimento organizacional e apoiando processos decisórios.

3 TRAJETÓRIA METODOLÓGICA

Neste capítulo é apresentada a trajetória metodológica utilizada, que buscou responder à questão de pesquisa e atender o objetivo geral e os específicos que nortearam o trabalho de pesquisa, identificando o tipo de abordagem, a natureza, os objetivos de pesquisa e os procedimentos.

3.1 MÉTODO

A pesquisa possui natureza aplicada, uma vez que tem como propósito gerar soluções práticas que contribuam com a melhoria dos processos de gestão e tomada de decisão de uma fundação de apoio, possibilitando impacto organizacional direto.

Quanto a abordagem, de caráter quantitativo, pois “busca responder a questões de extensão e tamanho de um fenômeno social, etc.” (Bufrem, Alves, 2020 p. 54).

Segundo Gil (2010) a pesquisa pode ser definida como o procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos.

Comungam também dessa ideia Lakatos e Marconi (2001) que definem como uma atividade voltada à busca de respostas e à solução de problemas para questões propostas, através da utilização de métodos científicos.

Quanto aos objetivos, caracteriza-se como exploratória e descritiva. Exploratória porque busca aprofundar a compreensão sobre o uso dos dados operacionais para suporte à decisão, identificando variáveis-chave e possíveis padrões ocultos na base analisada.

Descritiva porque apresenta, organiza e interpreta indicadores e relações entre variáveis do contexto real de gestão de projetos evidenciando o comportamento observado na base de dados estudada.

Segundo Bufrem e Alves (2020, p. 59-60), “estuda características de um grupo, sua distribuição, condições, opiniões ou qualidades que tenham significado para que se atinjam os objetivos da pesquisa”.

Como procedimento técnico, adota-se o estudo de caso único, desenvolvido em uma fundação de apoio vinculada a uma Instituição Federal de Ensino Superior

(IFES). A análise se concentra nos dados operacionais registrados no ERP Protheus no período de 2019 a 2024.

3.2 LEVANTAMENTO BIBLIOGRÁFICO

Para responder à pergunta dessa pesquisa: **a aplicação estruturada de técnicas de análise exploratória e mineração de dados podem apoiar o processo de tomada de decisão nas fundações de apoio?**

Inicialmente, foram definidos descritores padronizados com base nos cabeçalhos de assunto da área de Gestão da Informação, conforme o Tesauro Brasileiro da Ciência da Informação do IBICT.

As palavras foram traduzidas e combinadas da seguinte forma: *("organizational decision making process" OR "strategic decision making") AND ("business intelligence and decision making" OR "business intelligence" OR "data mining for business analytics" OR "data mining" AND "support foundation")*.

A pesquisa inicial, que incluía o termo *support foundation* (fundação de apoio), não resultou em nenhum artigo, indicando uma lacuna de estudos específicos sobre o uso de análise exploratória e mineração de dados no contexto das fundações de apoio.

Diante desse resultado, foi conduzida uma nova busca bibliográfica, utilizando os mesmos descritores, mas excluindo o termo *"support foundation"*. Essa segunda etapa teve o propósito de ampliar o escopo de pesquisa e identificar estudos relacionados à aplicação de técnicas de análise e mineração de dados em contextos organizacionais mais amplos, mas que pudessem ser adaptados à realidade das fundações de apoio.

Esses descritores foram combinados utilizando operadores booleanos, onde *"AND"* indica a associação de um termo com outro, e *"OR"* permite a busca de termos alternativos.

Uma estratégia inicial de busca foi elaborada para testar a aderência desses descritores nas bases de dados selecionadas. A escolha das bases considerou critérios de indexação em periódicos reconhecidos nacional e internacionalmente, assegurando a qualidade e relevância dos artigos identificados.

Após a definição dessa estratégica, foi realizada uma busca abrangente nas bases de dados da área de Ciências Sociais Aplicadas, disponíveis no Portal da CAPES. (<https://www-periodicos-capes-gov-br.ez22.periodicos.capes.gov.br/>).

As bases consultadas foram *EBSCOhost (Academic Search Ultimate, Business Source Premier, Computers & Applied Sciences Complete, Library, Information, Science & Technology Abstracts)*, *SCOPUS (Elsevier)*, e *Clarivate Analytics / Thomson Reuters - Web of Science – Collection*.

A pesquisa foi conduzida em inglês, de forma simultânea em todas as bases, no dia 10 de setembro de 2024, sendo restrita a artigos publicados nos últimos dez anos (2014-2024), revisados por pares e com busca nos campos de título, resumo e palavras-chave.

Além da estratégica de busca sistematizada, foram consultados autores renomados na área, selecionados a partir da frequência de citações em artigos científicos e sua contribuição para o campo pesquisado, caracterizando uma abordagem mista de levantamento (sistemático e exploratório).

As referências coletadas foram organizadas no gerenciador Zotero (<https://www.zotero.org/>). Os artigos relevantes foram selecionados com base na leitura dos títulos, resumos e palavras-chave, de acordo com os objetivos da pesquisa.

Abaixo pode ser observado no Quadro 3 que apresenta o Panorama da estratégia de busca.

QUADRO 3 – PANORAMA DA ESTRATÉGIA DE BUSCA

Base de dados	Período	Resultados 10/09/2024
<i>EBSCO - Academic Search Ultimate</i>	2013 a 2023	87
<i>EBSCO - Business Source Premier</i>	2013 a 2023	8
<i>EBSCO - Computers & Applied Sciences Complete</i>	2013 a 2023	85
<i>EBSCO - Library, Information Science & Technology Abstracts</i>	2013 a 2023	12
<i>Elsevier – Scopus</i>	2013 a 2023	45
<i>Clarivate Analytics / Thomson Reuters - Web of Science – Collection</i>	2013 a 2023	80

Fonte: O autor (2024).

Após a execução da estratégia de busca, identificou-se um total de 317 artigos revisados por pares, destes 30 foram excluídos pois estavam duplicados. Em seguida, foi realizada a leitura dos títulos e palavras-chave desses artigos com o intuito de verificar sua aderência ao tema de pesquisa.

Após a leitura dos títulos e palavras-chave 39 títulos demonstraram relevância e foram selecionados para a etapa seguinte de leitura dos resumos.

Os seguintes critérios foram adotados para a seleção dos artigos:

- Critérios de inclusão:
 - Artigos cujo título, resumo e palavras-chave apresentaram relação direta com o tema de pesquisa.
 - Artigos publicados em periódicos e revisados por pares.
 - Artigos disponíveis em acesso aberto ou com acesso viável a análise.
- Critérios de exclusão:
 - Artigos que não abordavam o tema principal da pesquisa ou tinham relação indireta com os objetivos.
 - Artigos com informações insuficientes para análise preliminar.
 - Artigos disponíveis sem acesso aberto.

Após a realização dos critérios de inclusão e exclusão restaram apenas 13 artigos, sendo esses utilizados como base para fundamentar a pesquisa segue abaixo o Quadro 4 com a apresentação desses artigos:

QUADRO 4 – ARTIGOS SELECIONADOS

Autor (es)	Título	Base	Ano
López-Robles, J.R; Otegi-Olaso, J.R; Porto Gómez, I. e Cobo, M.J.	<i>30 years of intelligence models in management and business: A bibliometric review.</i>	EBSCOhost	2019
Zhang, NC.	<i>A Campus Big-Data Platform Architecture for Data Mining and Business Intelligence in Education Institutes</i>	Web of Science	2016
Peral, J.; Maté. A. e Marco, M.	<i>Application of Data Mining techniques to identify relevant Key Performance Indicators.</i>	EBSCOhost	2017
Kudyba, S.	<i>Big Data, Mining, and Analytics: Components of Strategic Decision Making</i>	Scopus	2014

Moussas, K., Hafiane, J. e Achaba, A.	<i>Business intelligence and its pivotal role in organizational performance: An exhaustive literature review</i>	Scopus	2024
Alzghoul, A.; Khaddam, A. A.; Abousweilem, F.; Irtameh, H. J. e Alshaar, Q.	<i>How business intelligence capability impacts decision-making speed, comprehensiveness, and firm performance.</i>	EBSCOhost	2024
Murray, P. W.; Agard, B. e Barajas, M.A.	<i>Market segmentation through data mining: A method to extract behaviors from a noisy data set.</i>	Scopus	2017
Huang, S. Yue, H.	<i>Optimizing Strategic Decision-Making in Multinational Corporations through Data Mining-Based User Behavior Analysis.</i>	Scopus	2024
Arguir, D. C.; Dusa, I.G. e Onută, M.	<i>Organizational development through Business Intelligence and Data Mining.</i>	EBSCOhost	2019
Retnowardhani, A.; Sardjono, W. e Triana, Y.S.	<i>Review Study of Business Intelligence to Support Strategic Decision Making.</i>	Scopus	2019
Abu-Alsondos, I.A.	<i>The impact of business intelligence system (BIS) on quality of strategic decision-making.</i>	Scopus	2023
Serbanescu, L e Taicu, M.	<i>Using Business Intelligence Solutions for Company Performance Management.</i>	EBSCOhost	2015
West, S e Powell, D.	<i>Decision Making in Service Shops Supported by Mining Enterprise Resource Planning Data.</i>	EBSCOhost	2024

Fonte: O Autor (2025).

Além disso, foram consultadas partes de livros e artigos com o objetivo de complementar o levantamento bibliográfico e fornecer uma base teórica mais robusta

conforme apresentado abaixo no Quadro 5 o levantamento bibliográfico complementar:

QUADRO 5 – LEVANTAMENTO BIBLIOGRÁFICO COMPLEMENTAR

Autor (es), ano	Título	Tipo	Periódico ou Editora
Abreu (2024)	O papel crucial das fundações de apoios às IFES e ICTs: potencialidades e contribuições para o avanço científico e tecnológico.	artigo	Revista Tópicos
Barata e Prado (2015)	Governança de dados em organizações brasileiras.	artigo	SBC
Caetano e Campos (2019)	A autonomia das universidades federais na execução das receitas próprias.	artigo	Revista Brasileira de Educação
Castro e Ferrari (2016)	Introdução à mineração de dados.	livro	Editora Saraiva
Choo (2003)	A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões.	livro	Editora Senac
Davenport e Prusak (2003)	Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual.	livro	Editora Campus
Di Pietro (1998)	Direito Administrativo.	livro	Editora Atlas
Drummond (2008)	Gestão do conhecimento em organizações.	livro	Editora Saraiva
Fayyad, Piatetsky-Shapiro e Smith (1996)	<i>Knowledge discovery and data mining: towards a unifying framework.</i>	artigo	<i>Conference on Knowledge Discovery and Data Mining</i>
Fernandes (1994)	Privado, porém público - O Terceiro Setor na América Latina.	livro	Editora Relume/Dumará
Ferreira, Miranda, Pinto, et al. (2021)	Preparação e Análise Exploratória de Dados.	e-book	Editora SAGAH
Frank, Hall e Witten (2016)	<i>The Weka Workbench</i>	e-book	<i>Morgan Kaufmann</i>
Goldschmidt (2015)	<i>Data mining: conceitos, técnicas, algoritmos, orientações e aplicações.</i>	livro	Editora Elsevier

Marques e Correia Neto (2020)	Tomada de decisões gerenciais com analítica de dados: aplicações práticas com Excel.	livro	Editora Alta Books
Mendonça, Machado e Pinheiro (2004)	Governança nas organizações do terceiro setor: considerações teóricas.	artigo	Revista de Administração
Paes (2017)	Fundações, Associações e Entidades de Interesse Social - Aspectos jurídicos, administrativos, contábeis, trabalhistas e tributários.	livro	Editora Forense
Ramos, Rodrigues, Silva e Oliveira (2020)	<i>CRISP-DM</i> : uma proposta de adaptação do Modelo <i>CRISP-DM</i> para mineração de dados educacionais.	artigo	SBIE
Razzolini Filho (2009)	Gestão da informação e competências necessárias ao gestor.	livro	Editora SBGC
Rezende e Abreu (2013)	Tecnologia da Informação Aplicada a Sistemas de Informação Empresariais.	livro	Editora Atlas
Shearer (2000)	<i>The CRISP-DM Model: the new blueprint for data mining.</i>	artigo	<i>Journal of Data warehousing</i>
Sicsú; Sarmartini e Barth (2023)	Técnicas de machine learning.	livro	Editora Edgard Blücher
Stoner e Freeman (1999)	Administração.	livro	Editora LTC
Takeuchi e Nonaka (2008)	Gestão do conhecimento.	livro	Editora Bookman
Tarapanoff (2006)	Inteligência, informação e conhecimento em corporação.	livro	IBICT
Teixeira e Freitas Duarte (2014)	<i>Engineering information: conceptual elements related information management and information systems.</i>	artigo	<i>Edualearn14 Proceedings</i>
Yu (2011)	Tomada de decisão nas organizações: uma visão multidisciplinar.	livro	Editora Saraiva

Fonte: O autor (2024).

O levantamento bibliográfico deste projeto buscou identificar na literatura os achados e o que foi estudado e vem sendo estudado referente a análise exploratória de dados e sua aplicação na relação com as técnicas de mineração de dados, da gestão da informação e do conhecimento e da tomada de decisão, além de explicar melhor o tipo de organização que está sendo estudada.

Este estudo revelou a ausência de pesquisas previamente realizadas que integrem análise exploratória de dados, mineração de dados, gestão da informação e conhecimento e tomada de decisão no contexto de fundações de apoio. Isso indica que o tema ainda é pouco explorado na literatura e apresenta potencial para aprofundamento e contribuições significativas.

O levantamento apresentou três principais abordagens estatísticas: a Clássica (construção de modelo matemático antes da análise), Bayesiana (incorpora modelo probabilidade prévio, que pode ser atualizado dependendo de novos dados coletados) e a Análise Exploratória de Dados (ampla exploração inicial dos dados antes da definição do modelo estatísticos).

A literatura destaca que a AED se tornou uma etapa importante para organizações que pretendam tomar decisões baseadas em dados, pois busca examinar dados coletados e compreender as relações entre as variáveis antes de sua aplicação.

Outra abordagem encontrada na literatura trata-se da Descoberta de Conhecimento em Banco de Dados (KDD), que envolve fases de pré-processamento e preparação de dados (inclui limpeza, integração e transformação de dados), mineração de dados (aplicação de algoritmos para extração de padrões, utilizando técnicas como regressão, classificação, agrupamento (*clustering*), árvore de decisão) e interpretação dos resultados (análise dos padrões descobertos e garantir que sejam úteis para a tomada de decisão).

A literatura revisada ainda aponta a gestão da informação e do conhecimento como fatores decisivos para a tomada de decisão. Para os autores consultados, destacam que o valor estratégico dos dados está diretamente relacionado à sua qualidade, confiabilidade, tempestividade e capacidade de gerar conhecimento aplicável.

Nesse sentido, decisões organizacionais mais assertivas dependem de sistemas de informação capazes não apenas de armazenar grandes volumes de dados, mas também de processá-los, integrá-los e interpretá-los de forma inteligente.

Em ambientes complexos, como o de uma fundação de apoio, onde coexistem múltiplas fontes de dados operacionais como projetos, convênios, contratos, recursos e dispêndios, a ausência de mecanismos analíticos adequados pode comprometer a identificação de padrões e oportunidades estratégicas.

Assim, o desafio não está apenas em coletar informações, mas em transformá-las em conhecimento útil que subsidie o planejamento, a gestão de recursos e a sustentabilidade institucional.

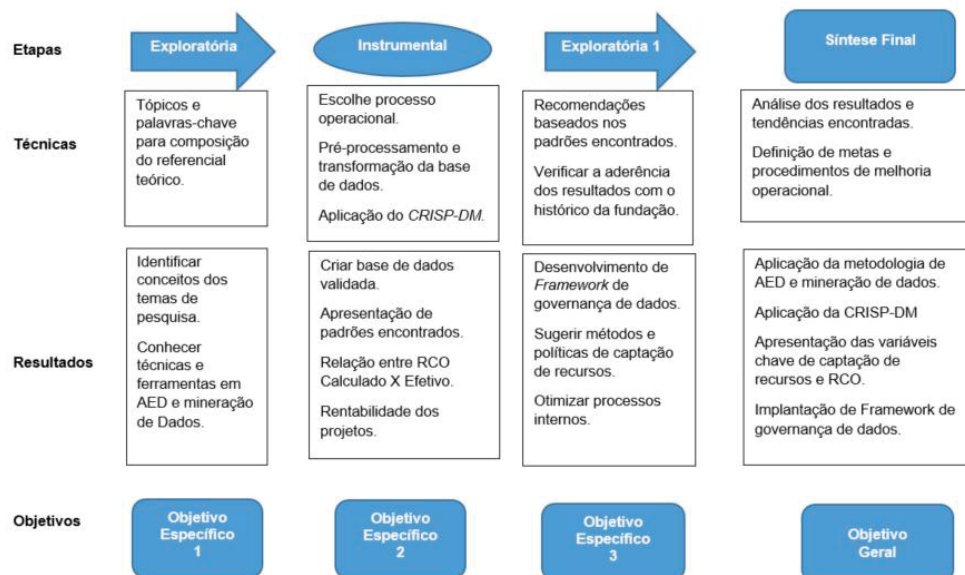
E o modelo SECI de gestão do conhecimento, apresenta que o processo de conhecimento nas organizações acontece por meio da conversão do conhecimento tácito e explícito, auxiliando para a inovação e competitividade organizacional.

4 PROPOSTA METODOLÓGICA

Com base nas referências teóricas apresentadas, as próximas etapas envolvem a seleção e o tratamento dos dados, aplicação de técnicas de análise exploratória, a validação dos padrões identificados e a apresentação dos resultados e padrões encontrados que possam subsidiar, de forma prática, a tomada de decisão pela direção da fundação

Em continuidade, a Figura 3 apresenta o fluxo metodológico proposto, detalhando os procedimentos que serão adotados para o alcance de cada objetivo geral e específico da pesquisa.

FIGURA 3 - SÍNTESE DAS ETAPAS DE PESQUISA



Fonte: O autor (2025).

4.1 MODELO DE REFERÊNCIA PARA O PROCESSO ANÁLITICO: CRISP-DM

Para estruturar o processo de extração de conhecimento a partir dos dados operacionais da Fundação de Apoio, esta pesquisa adota como base metodológica o modelo *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) (Shearer, 2000), reconhecido na literatura como referência para projetos de mineração de dados.

O CRISP-DM organiza o processo analítico em seis fases iterativas e interdependentes: compreensão do negócio, compreensão dos dados, preparação dos dados, modelagem, avaliação e implantação.

Sua flexibilidade permite adaptá-la a diferentes contextos institucionais, como é o caso desta pesquisa, voltada à análise de dados para suporte à tomada de decisão estratégica de uma Fundação de Apoio.

Embora o modelo CRISP-DM tenha sido adotado como estrutura metodológica principal, ele se alinha conceitualmente ao processo de descoberta de conhecimento em bases de dados (KDD) proposto por Fayyad *et al.* (1996), do qual deriva e que o fundamenta teoricamente, podendo ser compreendido com uma padronização prática do ciclo KDD.

Assim, a condução da pesquisa segue o modelo **CRISP-DM** (*Cross Industry Standard Process for Data Mining*), que estrutura o processo analítico em seis etapas: compreensão do negócio, compreensão dos dados, preparação, modelagem, avaliação e implantação. Cada uma dessas etapas está diretamente vinculada aos objetivos específicos da pesquisa, orientando a extração de conhecimento relevante para o suporte gerencial (Quadro 6).

QUADRO 6 - ETAPAS CRISP-DM

ETAPA	DESCRIÇÃO	APLICAÇÃO NO PROJETO
1. Compreensão do negócio	Entendimento do contexto institucional e dos objetivos estratégicos.	Mapear os principais desafios da Fundação, como gestão de contratos, captação de recursos e sustentabilidade financeira.
2. Compreensão dos dados	Coleta e análise inicial dos dados disponíveis para verificar sua relevância e qualidade.	Explorar os dados operacionais do Protheus para compreender a estrutura, tipo e disponibilidade da informação.
3. Preparação dos dados	Limpeza, transformação e formatação dos dados para análise posterior.	Realizar Análise Exploratória de Dados (AED), tratar dados ausentes, padronizar formatos e selecionar variáveis relevantes.
4. Modelagem	Aplicação de técnicas de mineração de dados, como clusterização, classificação ou associação.	Aplicar modelos para identificar padrões nos projetos, contratos e desempenho institucional da Fundação.
5. Avaliação	Verificação da qualidade e utilidade dos modelos construídos em relação aos objetivos do negócio.	Avaliar se os padrões identificados contribuem para melhorar a tomada de decisão na gestão da Fundação.
6. Implantação	Uso dos resultados obtidos para gerar valor à organização por meio de relatórios e ações.	Gerar relatórios, <i>dashboards</i> e recomendações estratégicas para apoiar a diretoria nas decisões institucionais.

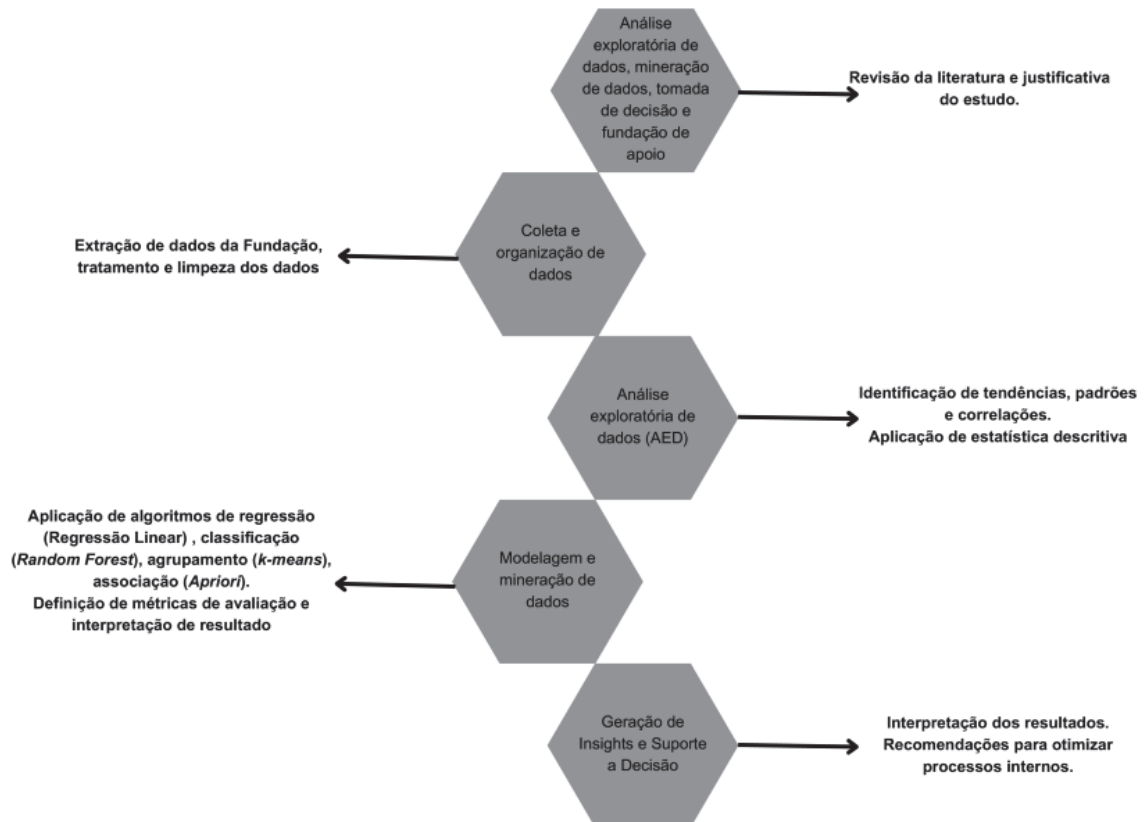
Fonte: O autor (2025).

As fases descritas no Quadro 6 são referência para a condução da pesquisa, estruturando a forma como os dados serão compreendidos, preparados, analisados e utilizados para apoiar a gestão estratégica da Fundação.

4.2 ETAPAS DA ANÁLISE EXPLORATÓRIA DE DADOS (AED)

Segue na Figura 4 abaixo que apresenta como serão realizadas as etapas da Análise Exploratória de Dados (AED).

FIGURA 4 - ETAPAS DA ANÁLISE EXPLORATÓRIA DE DADOS (AED)



Fonte: O autor (2025).

Além disso, será conduzida uma análise exploratória dos dados disponíveis no sistema de gestão de uma fundação de apoio vinculada à Universidade Federal do Paraná (UFPR).

A base de dados utilizada nesta pesquisa será composta por informações operacionais da fundação de apoio à UFPR, abrangendo aspectos relacionados à gestão administrativa e financeira de projetos.

Dentre os diversos processos operacionais conduzidos pela fundação optou-se por focar na análise do processo de captação de novos projetos, considerando a viabilidade temporal da pesquisa e a relevância estratégica desse processo para a instituição.

A escolha desse processo justifica-se pelo fato de que a captação de projetos está diretamente ligada com a missão da fundação, que é de “promover e realizar ações para o desenvolvimento na área de ensino, pesquisa e extensão, em apoio à UFPR e à sociedade”, bem como sua principal atividade, que envolve a “prospecção, elaboração e gestão de projetos” (Fundação, 2024).

A análise exploratória de dados relacionados a esse processo permitirá identificar padrões e tendências que possam contribuir para a otimização das estratégias de captação de recursos, fortalecimento e sustentabilidade institucional.

Para garantir a sustentabilidade financeira da fundação, torna-se fundamental a prestação de serviços de gestão administrativa e financeira de projetos de ensino, pesquisa e extensão, com a devida remuneração por essas atividades.

Cabe uma explicação importante: as fundações de apoio são remuneradas pela gestão administrativa e financeira que exercem em projetos de ensino, pesquisa, extensão, inovação e desenvolvimento institucional e tecnológico das Instituições Federais de Ensino Superior (IFES) ou aos Institutos de Ciência e Tecnologia (ICTs) às quais estão credenciadas na forma da Lei nº 8.958/94 para a prestação de serviços como fundação de apoio.

Contudo, para que essa remuneração ocorra, é necessário que as fundações comprovem os custos efetivamente incorridos na execução do projeto, ou seja, demonstrem o custo interno real da instituição para realizar a gestão do projeto.

Com base nisso, o Tribunal de Contas da União (TCU) determinou que as IFES exijam das fundações credenciadas uma metodologia capaz de mensurar e comprovar esses custos, que ficou denominada de ressarcimento pelas Despesas Operacionais Administrativas (DOA).

A fundação de apoio analisada, a partir desse direcionamento do TCU às IFES apoiadas, realizou um levantamento detalhado de suas atividades, processos, equipe e estrutura de custos diretos e indiretos, desenvolvendo uma metodologia própria para precificar cada uma das ações executadas.

Essa metodologia foi aprovada tanto pela IFES à qual a fundação está vinculada quanto pela Controladoria Geral da União (CGU). Mesmo após a aprovação dessa sistemática de ressarcimento, a remuneração da fundação de apoio enfrenta diversos problemas baseados em restrições legais.

A Resolução nº 41/2017 em seu artigo 34 diz que “...poderão ser ressarcidas, mediante a apresentação de memória de cálculo do rateio da despesa administrativa,

vedada a duplicidade ou a sobreposição de fontes de recursos no custeio de uma mesma parcela da despesa” (UFPR, 2017).

O parágrafo 1º deste mesmo artigo afirma que “O limite destinado ao ressarcimento dos custos administrativos da Fundação de Apoio, deverá observar limitações de entidade financiadora ou de legislação específica” (UFPR, 2017).

A limitação da remuneração da fundação está vinculada a no máximo 15% do valor total do projeto, percentual esse validado e aprovado pelo TCU, sendo que muitos editais impõem tetos ainda mais baixos, chegando a limitar essa remuneração a menos de 5%.

Diante desse cenário, torna-se fundamental ampliar a captação de recursos e reduzir a diferença entre o custo operacional calculado e o custo efetivo recuperado pela fundação na gestão dos projetos, além de analisar e rever os fluxos operacionais executados pela fundação de forma a otimizar e reduzir custos na execução de suas atividades internas.

As fontes de dados utilizadas nesta pesquisa serão os sistemas em operação na fundação, com destaque para o sistema Protheus da TOTVS, que fornecerá informações provenientes dos módulos de gestão de projetos, contratos, financeiro e contabilidade, bem como os bancos de dados internos da instituição.

Cabe destacar que a análise abrange os dados dos últimos seis anos (01/01/2019 à 31/12/2024). Esse período foi selecionado devido à recente transição do sistema de gestão de projetos interno para um sistema ERP (*Enterprise Resource Planning*), o que demandou um processo de adaptação das unidades e reestruturação dos fluxos operacionais.

Segundo a gestão da fundação, a partir desse período, os dados passaram a estar mais consolidados e normalizados, garantindo maior confiabilidade para a análise.

A metodologia adotada envolverá técnicas de mineração de dados, com ênfase e utilização da CRISP-DM e na análise exploratória de dados (AED) para identificar padrões e *outliers*.

Foram aplicados algoritmos de classificação, regressão, agrupamento e associação, com o objetivo de mapear as relações entre os coordenadores, contratos, setores e fontes de financiamento.

Essa abordagem visa identificar tendências, estabelecer metas e definir regras de negócios que possam otimizar a captação de recursos para projetos acadêmicos e institucionais.

Os resultados obtidos serão utilizados no planejamento estratégico da fundação nos próximos anos, contribuindo para o fortalecimento da gestão e ampliação das oportunidades de financiamento.

A análise de dados será conduzida por meio de extração, tratamento e visualização das informações. As principais ferramentas utilizadas incluem:

- a) *SQL Server*, para consultas diretas no banco de dados do sistema de gestão;
- b) *Python*, com bibliotecas como *Pandas*, *Scikit-Learn*, *Matplotlib* e *Seaborn*, para extração e modelagem de dados;
- c) *Power BI*, para criação de painéis interativos e visualizações dinâmicas das percepções identificadas;
- d) *Weka*, para aplicação de algoritmos de regressão, classificação, agrupamento (*clustering*) e associação, facilitando a descoberta de padrões e tendências nos dados.

A escolha dessas ferramentas justifica-se pela compatibilidade com os sistemas da fundação, suporte à integração com bancos *SQL*, capacidade de processamento de grandes volumes de dados e eficiência na geração de relatórios e visualizações estratégicas.

A experimentação foi realizada com o aplicativo *Weka* (Frank; Hall; Witten, 2016), aplicando-se diferentes modelos preditivos e técnicas de mineração de dados, permitindo uma abordagem comparativa para a melhor interpretação dos resultados.

5 APLICAÇÕES

Este capítulo apresenta a transição da exploração teórica e metodológica para a demonstração prática das contribuições desta dissertação. Com base nos fundamentos estabelecidos, especialmente da metodologia *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*), apresento a aplicação concreta dos modelos, ferramentas e análises realizadas.

Segundo Chapman (2000, citado por Ramos *et al.*, 2020), a metodologia, inclui descrições das fases normais de um projeto, as tarefas requeridas em cada fase e uma explicação sobre as relações entre as tarefas. Como modelo de processo, o *CRISP-DM* fornece uma visão geral do ciclo de vida da mineração de dados.

O objetivo central é ilustrar como a estrutura e as fases do *CRISP-DM* foram adaptadas e implementadas para abordar e auxiliar a solucionar o problema de pesquisa: **A aplicação estruturada de técnicas de análise exploratória e mineração de dados podem apoiar o processo de tomada de decisão nas fundações de apoio?**

Aplicando-se a metodologia baseada no *CRISP-DM*, observa-se abaixo as fases que serão exploradas e aplicadas à base de dados da fundação.

5.1 COMPREENSÃO DO NEGÓCIO

A Fundação de apoio realiza a gestão administrativa e financeira em projetos de ensino, pesquisa, extensão, desenvolvimento institucional e tecnológico para as Instituições Federais de Ensino Superior (IFES) às quais estão credenciadas para a prestação de serviços como fundação de apoio.

A fundação é remunerada através da DOA – Despesa Operacional Administrativa pela execução da gestão administrativa e financeira em cada um dos projetos executados.

Apresenta uma metodologia DOA de recuperação de custos operacionais baseada em seus custos efetivos diretos e indiretos e vinculadas a suas atividades, processos, equipe e estrutura.

O processo de construção da metodologia aconteceu em 2017, utilizando-se como base todos os projetos executados pela fundação nos anos anteriores. Sendo revisado em 2020.

Analisou-se os projetos que foram executados e o que o projeto utilizava da estrutura da fundação, com base em sua vigência, os recursos movimentados e os serviços prestados pela fundação.

Como ilustração das atividades realizadas destacam-se, a captação de recursos e novos projetos, o acompanhamento e apoio na negociação e formalização do projeto, o acompanhamento de sua execução, através da aquisição de bens, insumos e serviços, contratação de prestação de serviços de pessoa jurídica, contratação e pagamento de serviços prestados por autônomos, bolsistas (discentes, docentes), funcionários CLT, controle de plano de trabalho, fluxo financeiro e orçamentário, metas e atividades do projeto, transferência de bens até o seu prestações de contas parciais, anuais e finais até o encerramento e inclusive no pós projeto para atendimento de diligências que possam ocorrer.

Após a análise realizada, identificou-se alguns tipos de projetos executados conforme a origem do recurso e financiador e foram classificados também por níveis de porte conforme o montante de recursos previstos movimentados.

Ao final desse trabalho foi possível criar um sistema de precificação de cada uma das atividades realizadas pela fundação, por tipo de projeto e conforme a sua complexidade e demandas internas na fundação.

O objetivo desse estudo está em atender as obrigações legais que determinam que a remuneração da fundação deve estar vinculada a um modelo que seja possível a comprovação das despesas e que possa ser auditado.

Anualmente a metodologia de recuperação dos custos operacionais é atualizada com base nas informações das despesas diretas e indiretas da fundação nas Demonstrações Financeiras e Contábeis da fundação.

Esses demonstrativos são auditados por empresa de auditoria externa, aprovadas pelo Conselho Fiscal e Conselho Diretor da fundação de apoio, além de ser ratificado em conjunto com o Relatório de Atividades pela apoiada e pela prestação de contas anual ao Ministério Público das Fundações do Estado do Paraná.

Além disso, os valores das atividades são atualizados também conforme o orçamento previsto para o próximo exercício contábil com as despesas previstas da fundação.

Essa metodologia gera um mapa com as atividades disponíveis fim e meio e seu valor unitário e conforme o plano de trabalho do projeto e a utilização das atividades na fundação é gerado o valor da DOA.

Esses valores gerados pela metodologia não são estáticos, refletem diretamente os custos da fundação para a execução da gestão administrativa e financeira do projeto conforme o plano de trabalho, sendo possível ao coordenador do projeto analisar as necessidades de atividades e inclusive ao adequar o seu plano de trabalho reduzir o valor da DOA.

5.2 COMPREENSÃO DOS DADOS

Os projetos foram classificados após a análise da metodologia por tipo conforme abaixo:

- a) RCO SICONV: Projetos com a fundação de apoio à Universidade Federal do Paraná (UFPR), com captação de recursos pela conta única da UFPR e transferidos à fundação pelo SICONV (atual plataforma TransfereGov), com a utilização obrigatória para a execução do projeto por essa plataforma.
- b) RCO EDITAIS: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR), com captação de recursos pela conta única da UFPR e transferidos à fundação pelo SICONV, com a utilização obrigatória para a execução do projeto por essa plataforma. Compreende editais abertos pela universidade para abarcar o maior número possível de professores em um único projeto, são denominados na fundação como projetos com um único coordenador, mas com diversos professores contemplados.
- c) RCO TRIPARTITE: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR) financiados por terceiros, entes públicos ou privados com a finalidade de pesquisa.
- d) RCO CURSOS: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR), que compreendem cursos de especialização com captação por meio de mensalidades e/ou inscrições em conta específica do curso diretamente na fundação de apoio.
- e) RCO NÚCLEO DE CONCURSOS: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR), que compreende captação de recursos de concursos públicos, institucionais ou não, por meio de inscrições ou repasse de terceiros (pessoa jurídica pública ou privada), em conta específica do processo seletivo na fundação de apoio.

- f) RCO EVENTO: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR), que compreende projetos de curta duração com captação de recursos por meio de contratos de patrocínios e/ou inscrições em conta específica do evento na fundação de apoio.
- g) RCO CAPTACAO: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR), que compreende contrato entre a UFPR e a fundação para a gestão de recursos cobrados por serviços prestados por laboratórios e outros serviços da Universidade junto a terceiros em conta específica do projeto na fundação de apoio.
- h) RCO ESTUDOS CLINICOS: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR) e à Empresa Brasileira de Serviços Hospitalares (EBSERH), que compreende a captação por meio de repasse de terceiros (pessoa jurídica públicas ou privadas) nacionais ou internacionais para apoio a estudos e desenvolvimento de pesquisa clínica, em conta específica do projeto na fundação de apoio.
- i) RCO HOSPITAIS: Projetos como fundação de apoio à Universidade Federal do Paraná (UFPR), que compreende projetos hospitalares, onde a especificidade e complexidade dos serviços prestados pela fundação é maior, geralmente englobando a contratação de empregados CLT para o atendimento a demandas específicas da unidade hospitalar, podendo ser projetos institucionais com os hospitais universitários ou através de projetos Tripartite, onde os recursos serão executados em conta específica do projeto na fundação de apoio.

Cada um dos tipos de RCO identificados foram classificados com a inclusão da numeração de 1 a 5, conforme o valor do recurso previsto para o projeto, sendo classificado da seguinte forma:

- a) TIPO RCO 1: Projetos com recurso previsto até R\$100.000,00 (cem mil reais);
- b) TIPO RCO 2: Projetos com recurso previsto entre R\$100.000,01 (cem mil reais e um centavo) até R\$500.000,00 (quinhentos mil reais);
- c) TIPO RCO 3: Projetos com recurso previsto entre R\$500.000,01 (quinhentos mil reais e um centavo) até R\$1.000.000,00 (um milhão de reais);

- d) TIPO RCO 4: Projetos com recurso previsto entre R\$1.000.000,01 (um milhão de reais e um centavo) até R\$5.000.000,00 (cinco milhões de reais);
- e
- e) TIPO RCO 5: Projetos com recurso previsto acima de R\$5.000.000,01 (cinco milhões de reais e um centavo).

Algumas outras considerações devem ser apresentadas com relação aos dados da DOA – Despesas Operacionais Administrativas, principalmente no que tange a especificação de cada um dos tipos de RCO – Recuperação de Custos Operacionais apresentados no banco de dados utilizado para análise exploratória e mineração de dados.

RCO Calculado: É o valor da recuperação de custos operacionais da fundação de apoio com base na metodologia aprovada e no custo efetivo da fundação para executar o projeto (nesse momento o RCO Calculado pode estar acima das permissões legais de cobrança da DOA, pois trata-se do custo efetivo da fundação).

RCO Negociado: É o valor da recuperação de custos operacionais da fundação de apoio negociado para atender as determinações de limites legais ou de editais dos financiadores ou agências de fomento ou uma possível negociação com o coordenador do projeto para a execução do projeto pela fundação de apoio e registrado no instrumento assinado entre UFPR e fundação, seja ele convênio ou contrato.

RCO Efetivo: É o valor da recuperação de custos operacionais da fundação de apoio efetivamente recebido com base nos recursos recebidos e executados pelo projeto.

O objetivo principal da fundação para otimizar o ganho na recuperação de custos operacionais de cada projeto, é diminuir a diferença entre o RCO Calculado e o RCO Efetivo e garantir que o RCO Negociado fique o mais próximo do calculado e dentro das restrições legais.

Além disso, a fundação de apoio necessita reavaliar o fluxo dos processos e das atividades internas de modo a otimizar e melhorar a eficiência dos projetos atualmente realizados, através de automatização de processos hoje operacionais e manuais.

5.3 PREPARAÇÃO DOS DADOS

Nesse momento recorreremos à análise exploratória de dados para entender os dados disponíveis.

Com o intuito de buscar encontrar respostas para a melhora na captação de recursos e na recuperação dos custos operacionais da fundação de apoio, a base de dados apresentava projetos executados pela fundação com vigência a partir de 01/01/2019.

Após a análise inicial da base de dados completa ficou evidenciado que para uma melhor análise dos projetos executados com relação ao recurso previsto do projeto e a recuperação dos custos operacionais da fundação não seria coerente iniciar a análise de dados com projetos ainda vigentes, porque desta forma a execução ainda é parcial e não seria possível encontrar os principais gargalos com relação à remuneração da fundação.

Com base nisso, foram ajustadas a análise de dados para os projetos que já tinham a execução finalizada, tornando assim, mais coerente a análise.

Foram utilizados os projetos com vigência e encerramento no período de 01/01/2019 a 31/12/2024, que correspondem a períodos já encerrados e com as contas da fundação devidamente aprovadas pelo Conselho Fiscal, Conselho Diretor, a IFES apoiada e o Ministério Público das Fundações do Estado do Paraná.

Após esse ajuste na análise, a base de dados retornou 270 projetos, a base foi ajustada e teve algumas informações incluídas. A base utilizada apresenta os seguintes dados:

- a) FILIAL: Representa o número do projeto na fundação desde a sua abertura até o seu encerramento, cada projeto tem uma numeração única para o acompanhamento do cronograma físico e financeiro e garantir que não ocorra interferências entre os projetos como determina a legislação.
- b) DESCRICAO RCO: Representa a classificação do tipo de DOA – Despesas Operacionais Administrativas conforme o projeto que será executado pela fundação, baseado na metodologia aprovada.
- c) DATA INICIO VIGENCIA: Representa a data inicial de vigência do projeto conforme estabelecido no instrumento entre Fundação e IFES.
- d) DATA FIM VIGENCIA: Representa a data final da vigência do projeto conforme estabelecido no instrumento entre Fundação e IFES.

- e) VIGENCIA EM MESES: Representa o cálculo em meses da vigência do projeto, o período em que pode ser executado e ter movimentação financeira.
- f) TIPO DE VIGENCIA: Representa a classificação realizada pela fundação com relação a vigência do projeto conforme o tempo em meses para sua execução, sendo classificado da seguinte forma:
 - i Ultracurto: Execução do projeto em até seis meses;
 - ii Curto: Execução do projeto de seis meses e um dia a doze meses;
 - iii Médio: Execução do projeto de doze meses e um dia a vinte e quatro meses;
 - iv Longo: Execução do projeto de vinte e quatro e um dia a trinta e seis meses;
 - v Ultralongo: Execução do projeto de trinta e seis e um dia a quarenta e oito meses;
 - vi Estendido: Execução do projeto acima de quarenta e oito meses e um dia.
- g) UNIDADE GESTORA DA UFPR: Representa a unidade gestora responsável pelo projeto junto a IFES apoiada.
- h) FONTE DE RECURSOS: Refere-se ao tipo da fonte de recursos do projeto. Pode ser classificada da seguinte forma:
 - i UFPR: Recurso proveniente da Universidade Federal do Paraná;
 - ii Federal: Recurso proveniente de órgão da Administração Federal;
 - iii Estadual: Recurso proveniente de órgão da Administração Estadual;
 - iv Municipal: Recurso proveniente de órgão da Administração Municipal;
 - v Privado: Recurso proveniente de origem privada;
 - vi Serviços: Recurso proveniente de recebimento pela prestação de serviços;
 - vii Internacionais: Recurso proveniente de origem internacional;
 - viii Outros: Outros tipos de recursos não classificados anteriormente;
- i) DESCRICAO DE ATIVIDADE UFPR: Representa a vinculação do projeto com o tipo de atividade descrito pela UFPR, conforme a sua resolução e podem ser classificados em: Pesquisa, Ensino, Extensão, Desenvolvimento Institucional e Serviços Técnicos.

- j) **PORTE DO PROJETO:** Representa o porte do projeto classificado pela fundação conforme a previsão de recursos previstos para o projeto. São classificados da seguinte forma:
- i) **Micro:** Projetos com previsão de recursos até R\$300.000,00 (trezentos mil reais);
 - ii) **Pequeno:** Projetos com previsão de recursos de R\$300.000,01 (trezentos mil reais e um centavo) até R\$1.000.000,00 (um milhão de reais);
 - iii) **Médio:** Projetos com previsão de recursos de R\$1.000.000,01 (um milhão de reais e um centavo) até R\$5.000.000,00 (cinco milhões de reais);
 - iv) **Grande:** Projetos com previsão de recursos de R\$5.000.000,01 (cinco milhões de reais e um centavo) até R\$10.000.000,00 (dez milhões de reais);
 - v) **Excepcional:** Projetos com previsão de recursos acima de R\$10.000.000,01 (dez milhões de reais e um centavo).
- k) **PREVISAO DE RECURSO DO PROJETO:** Representa a previsão de recursos para o projeto conforme instrumento assinado.
- l) **RECURSO RECEBIDO PROJETO:** Representa o recurso efetivamente recebido pelo projeto.
- m) **% RECURSO RECEBIDO PROJETO:** Representa o percentual do recurso recebido pelo projeto em relação ao recurso previsto.
- n) **RENDIMENTO APLICACAO:** Representa os rendimentos de aplicação financeira auferidos pelo projeto durante sua vigência. O inciso IX, da Resolução 41/17 – COPLAD, determina que a fundação deverá: “Aplicar os recursos financeiros em conta poupança ou fundo de aplicação financeira de curto prazo ou operação de mercado aberto lastreada em título da dívida pública federal, quando não empregados na sua finalidade”.
- o) **% RENDIMENTO X RECURSO:** Representa o percentual do rendimento obtido em relação ao recurso recebido pelo projeto.
- p) **DISPÊNDIOS EXECUTADOS:** Representa os dispêndios executados pelo projeto incluindo as despesas efetivas do projeto, os pagamentos de taxas à Universidade Federal do Paraná (UFPR), a recuperação de custos operacionais da fundação de apoio e a devolução de recursos

remanescentes a Universidade Federal do Paraná (UFPR) quando for o caso.

- q) RCO CALCULADO: Representa o valor da recuperação de custos operacionais da fundação, calculada pelo método DOA aprovado da fundação.
- r) % RCO CALCULADO: Representa o percentual da recuperação de custos operacionais da fundação, calculada pelo método DOA aprovado da fundação.
- s) RCO NEGOCIADO: Representa o valor da recuperação de custos operacionais negociada e aprovada no instrumento entre fundação e IFES.
- t) % RCO NEGOCIADO: Representa o percentual da recuperação de custos operacionais negociada e aprovada no instrumento entre fundação e IFES.
- u) RCO EFETIVO: Representa o valor da recuperação de custos operacionais efetivamente recebidos pela fundação com base no recurso efetivamente recebido pelo projeto.
- v) % RCO EFETIVO: Representa o percentual da recuperação de custos operacionais efetivamente recebidos pela fundação com base no recurso efetivamente recebido pelo projeto.

Ao realizar a análise exploratória de dados na base de dados da fundação de apoio pesquisada, foi possível encontrar um sistema de apoio a gestão com diversas funcionalidades e os mais diversos campos disponíveis para a inserção de dados e informações relevantes.

No entanto, após uma análise mais apurada no banco de dados gerada através da consulta *SQL*, foi possível identificar muitas falhas operacionais, relacionadas ao preenchimento de diversos campos importantes e necessários no sistema.

Por isso, uma base de dados própria foi desenvolvida com as informações necessárias para atender ao projeto de pesquisa utilizando informações do Banco de Dados e incluindo os dados faltantes para garantir a confiabilidade dos dados para análise.

5.4 MODELAGEM

Foi realizada a análise exploratória de dados na base de dados dos 270 projetos executados e finalizados pela fundação no período 01/01/2019 a 31/12/2024.

A conformidade dos dados de uma base de dados a ser analisada é fato determinante para ser um suporte ou apoio a tomada de decisão.

5.4.1 Aplicação da Análise Exploratória de Dados

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), a análise exploratória de dados desempenha um papel essencial na descoberta de conhecimento, auxiliando gestores na extração de informações estratégicas para a otimização de processos.

A análise exploratória de dados busca examinar os dados coletados e compreender as relações entre as variáveis antes de sua aplicação em métodos estatísticos (Ferreira *et al.*, 2021).

Iniciou-se a análise da base de dados contendo os 270 projetos executados e finalizados pela fundação no período de 01/01/2019 a 31/12/2024, em apoio à UFPR, com o intuito de analisar a relação do valor recebido pela fundação a título de Recuperação de Custos Operacionais (RCO).

O Quadro 7, apresenta a relação entre o RCO EFETIVO e o RCO CALCULADO, apresentando o cálculo realizado pela metodologia da fundação quando da negociação e formalização do projeto.

QUADRO 7 – RELAÇÃO ENTRE RCO EFETIVO X RCO CALCULADO (em milhares de reais)

TIPO RCO	CONTAGEM FILIAL	SOMA RCO CALCULADO	SOMA RCO EFETIVO	RCO EFETIVO X RCO CALCULADO %
SICONV5	3	2.623	3.692	141
SICONV2	19	1.954	2.521	129
SICONV4	2	851	980	115
NUCLEO DE CONCURSOS4	8	1.950	1.936	99
EVENTOS3	1	106	102	96
NUCLEO DE CONCURSOS5	5	3.851	3.390	88
NUCLEO DE CONCURSOS2	5	361	291	81
SICONV3	4	378	288	76
TRIPARTITE3	11	620	443	71
CURSOS2	34	2.162	1.491	69
CAPTACAO2	6	230	142	62
CURSOS1	8	105	64	61
TRIPARTITE5	3	391	224	57
TRIPARTITE4	4	930	490	53
SICONV1	9	252	130	51
EVENTOS2	2	59	30	51
CURSOS3	40	2.974	1.467	49
CAPTACAO1	6	110	49	45

TRIPARTITE2	30	1.192	522	44
NUCLEO DE CONCURSOS3	2	208	87	42
NUCLEO DE CONCURSOS1	14	321	131	41
TRIPARTITE1	34	858	338	39
EVENTOS1	17	196	41	21
CAPTACAO5	2	1.184	152	13
ESTUDOS CLINICOS	1	38	1	2

Fonte: O autor (2025).

A base de dados analisada demonstra que 8,88% dos projetos executados pela fundação no período teve a sua Recuperação de Custos Operacionais (RCO), RCO EFETIVO, igual ou superior ao limite RCO CALCULADO no início do projeto.

Sendo esses projetos classificados como SICONV que representam recursos oriundos do orçamento próprio da Universidade Federal do Paraná (UFPR) ou repasses referentes às descentralizações orçamentárias de outros órgãos.

Nestes casos, observou-se que a execução do projeto e de seu recurso é realizado em sua totalidade, inclusive com a possibilidade de termos aditivos para inclusão de novos repasses de recursos e para sua execução.

Este tipo de projeto utiliza plataforma própria do governo federal (SICONV/TransfereGov) para a execução do projeto além de utilizar e registrar esses projetos no sistema de gestão da própria fundação, gerando dessa forma um retrabalho em toda a execução dos projetos desse tipo.

Outros 12,60% dos projetos executados o RCO EFETIVO realizou 70% do RCO CALCULADO, destaque para os projetos do tipo NUCLEO DE CONCURSOS, SICONV e TRIPARTITE.

Referem-se a projetos para execução de concursos públicos, convênios e contratos tripartite onde aparece a figura de um ente externo financiando o projeto em parceria com a Universidade Federal do Paraná (UFPR) e fundação.

O restante dos projetos, cerca de 78% dos projetos executados, o RCO EFETIVO realizou menos que de 70% do RCO CALCULADO, essa situação pode ser apontada em duas situações específicas: a primeira onde o projeto não tem sua execução realizada na totalidade fazendo com que a fundação não possa executar a sua recuperação na totalidade.

E a segunda, quando há restrições e limitações legais de cobranças da Recuperação dos Custos Operacionais (RCO) que em muitos editais específicos têm limitações podendo variar de 2,5% a 15%, nesses casos mesmo o RCO CALCULADO

ficando superior, a fundação por restrição legal pode ser remunerada somente até o limite permitido no edital.

Outro ponto que precisa de análise está na relação entre a execução dos recursos previstos e realizados nos projetos na busca identificar padrões e as principais diferenças nos recursos movimentados nos projetos.

O Quadro 8, apresenta essa relação entre o recurso que estava previsto a ser recebido no projeto e o recurso efetivo.

QUADRO 8 – RELAÇÃO ENTRE O RECURSO PREVISTO X RECURSO EFETIVO (em milhares de reais)

TIPO RCO	CONTAGEM FILIAL	SOMA PREVISAO RECURSO DO PROJETO	SOMA RECURSO RECEBIDO PROJETO	RECURSO EFETIVO X RECURSO PREVISTO %
EVENTOS2	2	275	325	118
NUCLEO DE CONCURSOS5	5	42.547	44.593	105
NUCLEO DE CONCURSOS4	8	17.007	17.520	103
NUCLEO DE CONCURSOS1	14	1.245	1.246	100
SICONV1	9	1.222	1.222	100
TRIPARTITE5	3	6.159	6.077	99
SICONV3	4	4.166	4.099	98
NUCLEO DE CONCURSOS2	5	2.581	2.507	97
EVENTOS3	1	1.250	1.200	96
SICONV5	3	33.514	32.038	96
TRIPARTITE4	4	10.614	9.744	92
SICONV2	19	37.324	33.271	89
TRIPARTITE1	34	3.828	3.400	89
SICONV4	2	10.093	8.915	88
CURSOS2	34	18.145	15.365	85
TRIPARTITE2	30	8.112	6.294	78
TRIPARTITE3	11	8.128	6.051	74
CAPTACAO1	6	549	372	68
CURSOS1	8	840	525	62
CURSOS3	40	21.720	13.415	62
CAPTACAO2	6	1.450	853	59
NUCLEO DE CONCURSOS3	2	1.494	872	58
EVENTOS1	17	1.111	601	54
ESTUDOSCLINICOS	1	21	8	36
CAPTACAO5	2	13.000	2.436	19

Fonte: O autor (2025).

Ao analisar a relação entre o RECURSO PREVISTO para o projeto e o RECURSO EFETIVO, verificou-se que 65% dos projetos executados receberam 75% ou mais do recurso que estava previsto e destes 32% projetos receberam mais de 90% do recurso previsto.

Isso representa uma execução efetiva muito boa para os projetos na execução da previsão dos recursos previstos, no entanto, a relação entre recurso recebido pelo projeto não pode ser constatada no RCO EFETIVO cobrado pela fundação.

Dos 188 projetos executados pela fundação que executaram 75% ou mais do RECURSO PREVISTO pelo projeto, 82 desses projetos são do tipo TRIPARTITE, projetos que envolvem a fundação, a UFPR e um outro financiador.

As submissões desse tipo de projetos são regidas por editais de empresas ou agências de fomento onde o RCO não pode ultrapassar 5% do valor do recurso executado pelo projeto.

Nesses casos, mesmo que o RCO CALCULADO seja superior, não será possível para a fundação obter a devida remuneração devido a restrições do edital e legais, apresentando um padrão e um risco para a fundação na execução desses tipos de projetos.

Outro ponto a ser destacado e pode ser evidenciado na análise que os projetos que executaram abaixo de 70% do RECURSO PREVISTO, em 80 dos casos referem-se a projetos de captação (cursos, evento, prestação de serviços (captação), etc.), que dependem diretamente da entrada do recurso para a execução do projeto.

Nesses projetos apresentou também um padrão que gera um ponto de atenção à fundação: como melhorar a captação desses projetos? Existem formas de melhorar a captação com um planejamento e apoio de comunicação e marketing? Qual é a carga desses projetos para a execução e prestação de serviços? Existe uma análise de mercado ou plano de negócio para esses tipos de projetos?

Esses são itens importantes a serem analisados para que a fundação possa auxiliar os coordenadores de projetos na execução administrativa e financeira e melhorar a sua captação de recursos e por consequência melhorar a remuneração da fundação.

O grande problema enfrentado pelas fundações de apoio está relacionado com a forma de metodologia de cobrança dos serviços prestados que são aprovadas e autorizadas pelos órgãos de controle.

A fundação não pode realizar a cobrança do preço dos seus serviços, deve cobrar o valor referente a recuperação de seus custos operacionais, ou seja, o custo que aquele projeto gerou para a fundação, mas limitado às questões dos percentuais legais e conforme limites dos financiadores.

Além disso, os projetos que não executam as receitas previstas têm a necessidade de readequar as suas despesas do plano de trabalho com base nos recursos efetivamente calculados, gerando a necessidade de diversos termos aditivos de supressão de valores e atividades o que gera a redução da DOA da fundação.

Assim, existe a necessidade de melhorar e reduzir o problema, representado pela alta diferença entre o RCO CALCULADO e o RCO EFETIVO, no entanto, isso torna-se inviável quando não há possibilidade legal de cobrança além dos limites autorizados.

Por outro lado, a fundação pode desenvolver e aplicar metodologias voltadas à estruturação de projetos de captação de recursos, cursos, eventos, como estratégias para fomentar a arrecadação.

A adoção de práticas de marketing, comunicação institucional e ferramentas de apoio à prestação de serviços pode contribuir significativamente para ampliar o alcance e fortalecer o relacionamento com o público-alvo.

Essas ações podem colaborar também para a redução da inadimplência, além de uma cobrança efetiva da fundação, o que por consequência, tende a aumentar os recursos efetivamente captados pelos projetos.

Com maior aderência às metas financeiras e operacionais, uma validação e um plano de negócio para a execução do plano de trabalho do projeto dentro das expectativas e baseadas na análise de mercado, torna-se possível garantir a plena execução das atividades previstas, assegurando também o recebimento do RCO EFETIVO pela fundação sempre respeitando os limites legais estabelecidos para sua remuneração.

Pensando nessa relação, a análise de dados continuou buscando encontrar uma relação entre os pontos já apresentados do RCO CALCULADO, RCO EFETIVO e o RECURSO PREVISTO E RECURSO EFETIVO.

De forma a mapear quais os tipos de projetos seriam mais rentáveis à fundação, levando em consideração somente essas duas relações, temos os seguintes resultados, conforme o QUADRO 9.

QUADRO 9 – RENTABILIDADE POR TIPO.

TIPO RCO	CONTAGEM DE FILIAL	RCO EFETIVO X RCO CALCULADO %	RECURSO EFETIVO X RECURSO PREVISTO %	RENTABILIDADE POR TIPO %
SICONV5	3	141	96	147
SICONV2	19	129	89	145

SICONV4	2	115	88	130
CAPTACAO2	6	62	59	105
EVENTOS3	1	96	96	100
CURSOS1	8	61	62	98
NUCLEO DE CONCURSOS4	8	99	103	96
TRIPARTITE3	11	71	74	96
NUCLEO DE CONCURSOS5	5	88	105	84
NUCLEO DE CONCURSOS2	5	81	97	83
CURSOS2	34	69	85	81
CURSOS3	40	49	62	80
SICONV3	4	76	98	77
NUCLEO DE CONCURSOS3	2	42	58	72
CAPTACAO5	2	13	19	68
CAPTACAO1	6	45	68	66
TRIPARTITE5	3	57	99	58
TRIPARTITE4	4	53	92	57
TRIPARTITE2	30	44	78	56
SICONV1	9	51	100	51
TRIPARTITE1	34	39	89	44
EVENTOS2	2	51	118	43
NUCLEO DE CONCURSOS1	14	41	100	41
EVENTOS1	17	21	54	39
ESTUDOS CLINICOS	1	2	36	7

Fonte: O autor (2025).

Ao buscar a relação entre o RCO CALCULADO, EFETIVO e o RECURSO PREVISTO e RECURSO Efetivo do projeto, calculou-se a Rentabilidade por tipo de projeto, com a utilização e a aplicação das seguintes equações:

1. % RCO REALIZADO

$$RCO \text{ Realizado} = \frac{RCO \text{ Efetivo}}{RCO \text{ Calculado}} * 100$$

2. % Receita Realizada

$$Receita \text{ Realizada} = \frac{Receita \text{ Efetiva}}{Receita \text{ Prevista}} * 100$$

3. Rentabilidade por tipo

$$Rentabilidade \text{ por tipo} = \frac{RCO \text{ Realizado}}{Receita \text{ Realizada}} * 100$$

Ao aplicar a equação para encontrar a rentabilidade por tipo de projeto identificou-se que os projetos do tipo SICONV (2, 4 e 5) e CAPTACAO (2) e EVENTOS

(3) foram os projetos que trouxeram uma melhor rentabilidade bruta na comparação das variáveis selecionadas.

Extraiu-se que apesar dos projetos do tipo SICONV serem os que tem sobreposições de atividades e duplicidades de tarefas para a fundação ainda são os mais rentáveis.

Esses projetos tendem a diminuir para os próximos anos, uma vez que foi autorizado às Instituições Federais de Ensino Superior (IFES) que realizem contratos diretamente com a sua fundação apoiada sem a necessidade de utilização do sistema.

Isso trará mais agilidade na execução do projeto e para a fundação, mas por outro lado uma redução no valor da recuperação de custos operacionais pela redução de atividades realizadas.

5.4.2 Aplicação da Mineração de Dados

Após a análise exploratória de dados foram aplicados alguns algoritmos de mineração de dados (classificação, regressão, clusterização e associação) através do *Python* para investigar a base de dados buscando encontrar padrões, fazer e gerar simulações que melhorem a captação de recursos para os projetos e a fundação.

A escolha dos algoritmos adotados nesta etapa foi orientada pela natureza dos dados e pelos objetivos de análise definidos. Buscou-se empregar diferentes abordagens: preditiva, explicativa e exploratória de modo a compreender o comportamento das variáveis financeiras e operacionais dos projetos sob múltiplas perspectivas.

O algoritmo de classificação *Random Forest* foi selecionado por sua robustez frente a bases heterogêneas e desbalanceadas, além oferecer uma interpretabilidade por meio da identificação das variáveis, mas relevantes na formação das classes.

A regressão linear foi aplicada para estimar relações quantitativas entre o RCO EFETIVO e as variáveis explicativas, permitindo avaliar o quanto as previsões do modelo se aproximam dos valores observados.

A escolha desse modelo deve-se à sua simplicidade interpretativa e transparências, características adequadas para aplicações exploratórias em contextos organizacionais.

Os modelos mais complexos, como regressão não linear ou redes neurais, foram descartados nesta etapa devido à ênfase em compreensibilidade e não apenas em desempenho preditivo.

O algoritmo *k-means* foi utilizado para agrupar os projetos com base em semelhanças de comportamento entre o RCO EFETIVO e a RECEITA REALIZADA. Esse método foi escolhido pela eficiência na segmentação de dados numéricos e pela facilidade de visualização dos resultados, características que o tornam adequado para análise de portfólios de projetos.

Outros métodos, como *DBSCAN* ou hierárquico, foram avaliados, mas não aplicados devido à ausência de padrões densos ou relações hierárquicas nos dados analisados.

Por fim, o algoritmo *Apriori* foi utilizado para explorar regras de associação entre variáveis categóricas, permitindo identificar possíveis combinações recorrentes que pudessem indicar padrões de comportamento.

Apesar de não terem sido encontradas associações estatisticamente relevantes, o uso do *Apriori* foi importante para confirmar a independência entre certas variáveis, reforçando a consistência dos resultados obtidos nas análises anteriores.

Dessa forma, a combinação desses algoritmos procurou equilibrar o rigor técnico, interpretabilidade e aplicabilidade prática, permitindo extrair conhecimento estratégico dos dados operacionais de maneira compreensível e útil para o processo de tomada de decisão institucional.

Segue abaixo o Quadro 10, com o comparativo dos algoritmos utilizados.

QUADRO 10 – COMPARATIVO DE ALGORITMOS APLICADOS.

Técnica	Tipo	Justificativa	Requisitos	Resultado Esperado
<i>Random Forest</i>	Classificação	Robustez com dados heterogêneos.	Dados numéricos categóricos.	Prever classe rentabilidade
Regressão Linear	Regressão	Estimar RCO EFETIVO.	Variáveis contínuas.	Predição de RCO.
<i>k-means</i>	Clusterização	Agrupar perfis de projetos.	Dados normalizados.	Identificar grupos de rentabilidade.
<i>Apriori</i>	Associação	Detectar padrões entre variáveis.	Dados categóricos.	Regras de associação entre

				RCO/Receita e desempenho.
--	--	--	--	---------------------------

Fonte: O Autor (2025).

Seguindo com as análises, a primeira etapa foi criar uma base de dados contendo os seguintes campos e informações: TIPO RCO, PREVISÃO RECURSO PROJETO, RECURSO REALIZADO PROJETO, RCO CALCULADO, RCO EFETIVO e foi criado um campo novo denominado CLASSE RENTABILIDADE (onde a classe da rentabilidade será classificada em Alta, Média e Baixa)

A Classe Rentabilidade será classificada da seguinte forma:

- Alta: quando o percentual do RCO EFETIVO for superior a 80% do RCO CALCULADO;
- Média: quando o percentual do RCO Efetivo estiver entre 60 e 79,99% do RCO Calculado;
- Baixa: quando o percentual do RCO Efetivo for inferior a 60% do RCO Calculado.

Ao utilizar o algoritmo de classificação *Random Forest* ele apresentou algumas constatações interessantes que podem ser verificadas no Quadro 11 a seguir, que apresenta os resultados do teste realizado.

QUADRO 11 – RESULTADO TESTE *RANDOM FOREST*

Classe	TP Rate (Recall)	FP Rate	Precision	F1-Score	ROC Area	PRC Area	Interpretação
ALTA	0.841	0.479	0.606	0.704	0.802	0.680	Modelo acerta bem essa classe.
BAIXA	0.481	0.159	0.662	0.557	0.581	0.398	Resultados medianos.
MÉDIA	0.105	0.060	0.222	0.143	0.526	0.142	Modelo quase não consegue reconhecer essa classe.

Fonte: O autor (2025).

O Quadro 11 apresenta uma avaliação pormenorizada do desempenho do classificador *Random Forest*, desagregada por classe, utilizando métricas cruciais para a análise de modelos de classificação, especialmente em cenários com potencial desbalanceamento de classes.

A classe ALTA demonstrou o desempenho mais robusto. O modelo alcançou uma Taxa de Verdadeiros Positivos (*TP Rate/Recall*) de 0,841, confirmando que 84,1% dos projetos pertencentes a essa classe foram identificados corretamente. Este valor elevado corrobora a hipótese de que os projetos da classe ALTA possuem atributos distintivos e bem definidos que o modelo é capaz de capturar.

Contudo, a Precisão (*Precision*) de 0,606 revela uma limitação: das previsões classificadas como ALTA, apenas 60,6% estavam corretas. Essa discrepância é reflexo do alto Taxa de Falsos Positivos (*FP Rate*) 0,479. Isso significa que, embora o modelo seja eficaz em encontrar os casos na classe ALTA (alto *Recall*), ele também comete um número significativo de erros de TIPO I, classificando erroneamente quase metade das instâncias negativas como ALTA.

O *F1-Score* de 0,704 reflete um bom equilíbrio entre a capacidade de cobertura (*Recall*) e a exatidão (*Precision*), sendo um desempenho satisfatório para a distinção desta classe. A Área *ROC* (0,802) confirma a boa capacidade discriminatória do modelo para a classe.

O desempenho para a classe BAIXA pode ser caracterizado como mediano, mas com um perfil de erro distinto, O *Recall* (0,481) é significativamente inferior, indicando que o modelo falou em identificar mais da metade dos projetos que verdadeiramente pertenciam a essa classe (elevada ocorrência de Falsos Negativos).

Em contrapartida, a Precisão (0,662) é mais alta entre as três classes. Este achado, complementado pelo baixo *FP Rate* (0,159), sugere que o modelo é altamente seletivo e confiável ao prever a classe BAIXA.

As previsões feitas são mais prováveis de estarem corretas, mas o custo é a omissão de uma grande parte dos casos reais (baixa cobertura). O *F1-Score* de 0,557 e a Área *ROC* (0,581) modestas confirmam a dificuldade do modelo em generalizar o padrão dessa classe.

A classe MÉDIA apresenta o maior desafio e a falha mais significativa do modelo. Com um *Recall* de apenas 0,105, o modelo mal conseguiu reconhecer a existência dessa classe. A dificuldade é acompanhada pela menor Precisão (0,222) e o *F1-Score* mais baixo (0,143), confirmando que a classificação para esta classe é pouco confiável e quase aleatória.

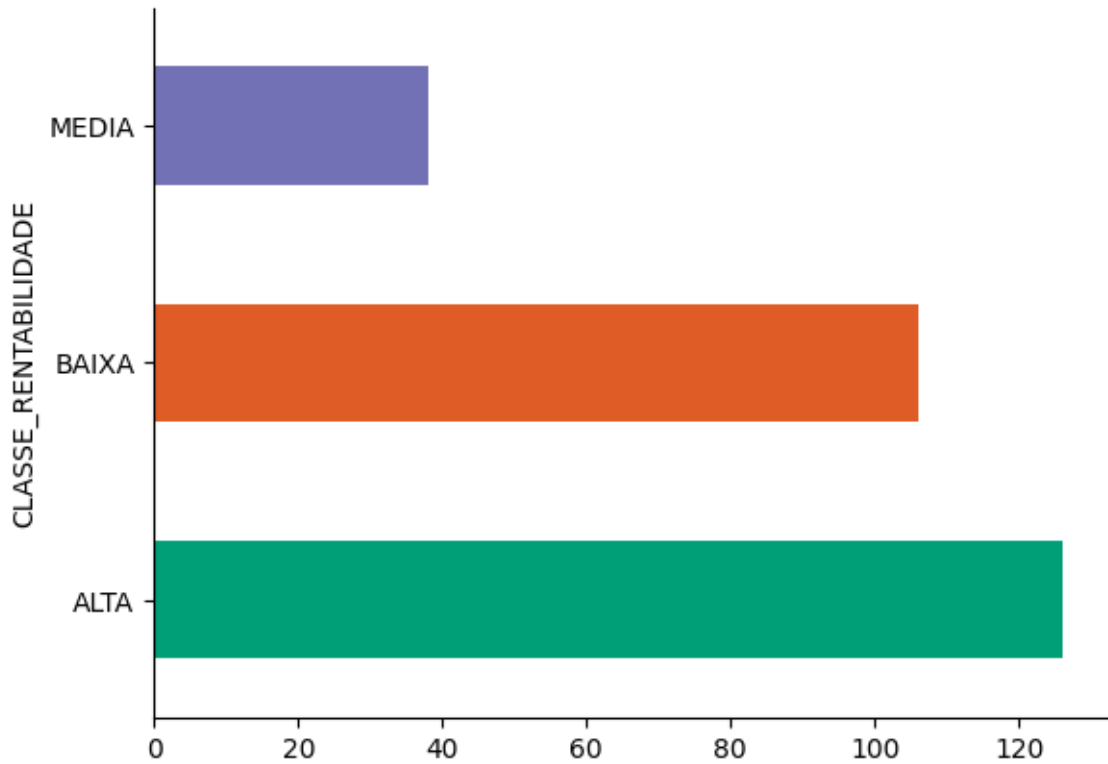
Embora o *FP Rate* (0,060) seja extremamente baixo, isso não é um indicador de bom desempenho, mas sim um reflexo de que o modelo quase nunca prevê a classe MÉDIA, optando por alocar a maioria das instâncias Média (Falsos Negativos) nas classes ALTA e BAIXA.

Os valores de Área *ROC* (0,526) e, notavelmente, a Área *PRC* (0,142), muito próximos a 0,5 e 0,0, respectivamente, demonstram que o modelo não consegue diferenciar efetivamente essa classe das demais.

Em sequência a Figura 5, abaixo apresenta a relação entre a Classe Rentabilidade, onde o eixo X representa o número de projetos em cada categoria de rentabilidade, enquanto o eixo Y indica as classes rentabilidade (ALTA, MÉDIA e BAIXA), definidas a partir da proporção entre o RCO EFETIVO e RCO CALCULADO.

Os preditores utilizados para classificar os projetos incluem o RCO CALCULADO, o RCO EFETIVO, a previsão de recurso do Projeto e o tipo do RCO. O gráfico evidencia a distribuição dos projetos entre as classes, permitindo observar que a maioria dos projetos apresenta rentabilidade ALTA, seguida da classe BAIXA, enquanto a classe MÉDIA apresentar menor rentabilidade.

FIGURA 5 – CLASSE RENTABILIDADE



Fonte: O autor (2025).

O QUADRO 12 detalha a performance do algoritmo de classificação *Random Forest* por meio da Matriz de Confusão e das métricas resultantes, fornecendo uma visão aprofundada dos acertos e erros do modelo em cada categoria.

QUADRO 12 – MATRIZ DE CONFUSÃO E MÉTRICAS DETALHADAS POR CLASSE (*RANDOM FOREST*)

Classe	TP	FP	FN	TN	Precision	Recall	F1-score	Support
Alta	25	6	13	37	0,8065	0,6579	0,7246	38
Baixa	23	9	9	40	0,8214	0,7188	0,7667	32

Média	4	7	7	63	0,1818	0,3636	0,2424	11
Macro Avg	-	-	-	-	0,6032	0,5801	0,5779	81
Weighted Avg	-	-	-	-	0,7275	0,642	0,6758	81

Fonte: O autor (2025).

Os resultados verificados confirmaram um bom desempenho nas classes majoritárias (ALTA e BAIXA), com um comprometimento notável na classe minoritária (MÉDIA).

A classe ALTA, apresenta uma *Precision* (0,8065) robusta, indicando que a maioria das previsões positivas para esta classe é correta. Contudo, o *Recall* (0,6579), embora aceitável, é inferior à *Precision*.

A ocorrência de 13 Falsos Negativos (*FN*) sugere que o modelo tende a ser conservador, falhando em identificar 13 instâncias, enquanto os 6 Falsos Positivos (*FP*) demonstram que poucas instâncias de outras classes foram indevidamente classificadas como ALTA. O *F1-Score* de 0,7246 é satisfatório.

A classe BAIXA, demonstra o desempenho mais equilibrado, atingindo a maior *Precision* (0,8214) e o maior *Recall* (0,7188). Com 9 Falsos Negativos (*FN*) e 9 Falsos Positivos (*FP*), a proporção de erros TIPO I e TIPO II é similar, resultando no maior *F1-Score* (0,7667), evidenciando uma alta capacidade de generalização e discriminação para esta categoria.

A classe MÉDIA, o desempenho é significativamente comprometido. O *Recall* (0,3636) indica que apenas 36,36% dos casos reais da classe foram detectados. A *Precision* (0,1818) é crítica, demonstrando que mais de 80% das previsões de MÉDIA estavam incorretas.

O *F1-Score* de 0,2424 atesta baixa capacidade do modelo em aprender e distinguir os padrões dessa classe minoritária, reforçando o impacto do desbalanceamento.

Em uma etapa subsequente, a análise preditiva foi estendida à previsão do RCO EFETIVO, utilizando a Regressão Linear, cujas métricas de erro foram as seguintes:

- Mean Squared Error (MSE)*: 0,0042.
- Root Mean Squared Error (RMSE)*: 0,065.
- Mean Absolute Error (MAE)*: 0,036
- R^2 : -0,03

O *RMSE* de 0,065 sugere que o desvio médio das previsões do modelo na escala do RCO EFETIVO, é de 6,5%. O *MAE* de 0,036 (3,6%) é inferior ao *RMSE*, indicando que, embora o erro médio absoluto seja baixo, a diferença entre as duas métricas sugere a presença de alguns *outliers* de erro, os quais o *RMSE* penaliza severamente.

No entanto, o coeficiente R^2 negativo (-0,03) é um achado crítico. Na literatura de *machine learning*, um R^2 negativo implica que o modelo linear testado é estatisticamente inferior a um modelo base que simplesmente prevê a média dos dados observados. Consequentemente, o modelo, com o conjunto de preditores utilizado, não conseguiu explicar a variabilidade da variável dependente RCO EFETIVO.

A análise comparativa entre cenários de balanceamento revela o seguinte trade-off fundamental:

- a) Com *class_weight*: O modelo priorizou as classes majoritárias (*Recall*: 0,82 e 0,84 em detrimento da classe minoritária (*Recall*: 0,18), mantendo a estrutura original dos dados.
- b) Com *SMOTE*: A aplicação da técnica de *oversampling* sintético aumentou o poder de detecção da classe minoritária (*Recall*: 0,36), mas essa melhoria ocorreu à custa da acurácia global (queda de 0,74 para 0,64) e da sensibilidade das classes majoritárias.

Esses resultados demonstram que as técnicas de balanceamento impactam de forma distinta a sensibilidade e a especificidade do modelo. O aumento do *Recall* na classe minoritária por meio do *SMOTE* introduziu ruído nos dados sintéticos, o que comprometeu o desempenho geral.

Em virtude das limitações temporais, estratégias de balanceamento avançadas (como *threshold tuning* e validação estratificada) não foram implementadas, mas a recomendação formal à gestão para a inclusão dessas técnicas em estudos futuros é validada pela análise.

Além disso, a baixa performance na Regressão Linear ($R^2 < 0$) e na classificação da classe MÉDIA apontam necessidade de explorar a inclusão de variáveis adicionais, transformações de dados mais sofisticadas ou a seleção de modelos preditivos alternativos para aumentar a capacidade explicativa e preditiva sobre o RCO.

Em continuidade à fase de modelagem, buscou-se uma compreensão mais granular dos padrões operacionais e financeiros dos projetos por meio da análise de *cluster*.

O objetivo central foi segmentar o portfólio da fundação, identificando grupos homogêneos (*clusters*) de projetos em função de duas variáveis críticas: a RECEITA REALIZADA do Projeto com o RCO EFETIVO.

Para este fim, aplicou-se o algoritmo *k-means*, utilizando $k = 3$ clusters. A escolha deste valor foi validada metodologicamente através do Coeficiente de Silhueta Médio conforme apresentado no Quadro 13 abaixo:

QUADRO 13 – COEFICIENTE DE SILHUETA MÉDIO POR NÚMERO DE *CLUSTERS* (k)

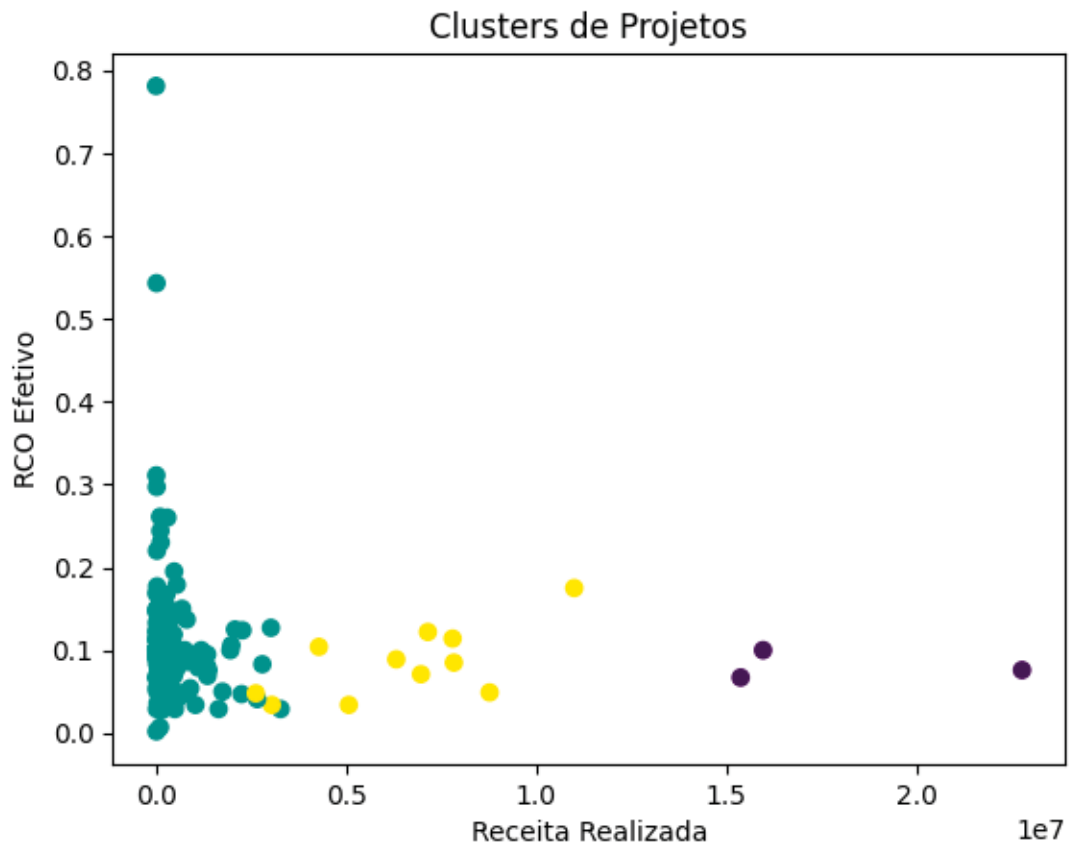
Valor de k	Coeficiente de Silhueta Médio
2	0.7988
3	0.7970
4	0.6164
5	0.5193
6	0.5265
7	0.5245
8	0.5625
9	0.4732
10	0.4933

Fonte: O Autor (2025).

A análise demonstrou que, embora $k = 2$ tenha alcançado o máximo estatístico (0,7988), o resultado para $k = 3$ (0,7970) é marginalmente inferior. Dada a necessidade gerencial de segregar o portfólio em três perfis distintos (pequeno, médio e grande porte de receita), a adoção de $k = 3$ é justificada, pois oferece a melhor combinação entre qualidade estatística (Alta Silhueta) e relevância estratégica. A robustez da escolha é reforçada pela queda acentuada na métrica a partir de $k = 4$ (0,6164), que invalida a segmentação.

Um número que mostrou eficaz na captura das principais diferenciações estruturais do portfólio. O resultado dessa segmentação é visualizado na Figura 6 (Clusters de Projetos: RCO EFETIVO x RECEITA REALIZADA), onde cada cluster (cor) representa um perfil de projeto distinto:

FIGURA 6 – CLUSTERS DE PROJETOS RCO EFETIVO x RECEITA REALIZADA



Fonte: O autor (2025).

Cada ponto representa um projeto e cada cor representa um *cluster*, observe-se as seguintes análises:

Cluster verde: projetos concentrados à esquerda.

- a) Receita realizada muito baixa (a maioria até R\$2 milhões de reais).
- b) RCO EFETIVO variado: de muito baixo até 80%.
- c) Esse grupo tem grande quantidade de projetos, com baixa receita, mas alguns com alta eficiência no RCO EFETIVO.

Cluster amarelo: disperso entre R\$2 e R\$10 milhões de reais.

- a) Receita realizada média (geralmente entre R\$2 e R\$10 milhões de reais).
- b) RCO EFETIVO entre 0% e 15%.
- c) Projetos que movimentam mais recursos, mas têm RCO EFETIVO mais controlado (não tão alto).
- d) Podem representar projetos grandes com margem menor.

Cluster Roxo: extremo direito.

- a) Receita realizada muito alta (acima de R\$10 milhões, até R\$20 milhões de reais ou mais).

- b) RCO EFETIVO baixo inferior a 10%.
- c) Provavelmente são projetos grandes, mas pouco rentáveis proporcionalmente.
- d) Importante analisar se o baixo RCO EFETIVO é esperado ou trata-se de projetos com restrições de recuperação limitas ao edital.

O Quadro 14, abaixo apresenta as estatísticas por *Cluster*.

QUADRO 14 – ESTATÍSTICAS POR *CLUSTER* X TICKET MÉDIO X RECEITA X QUANTIDADE DE PROJETOS

Cluster	Ticket Médio (R\$)	Desvio Receita	Qtd Projetos	RCO Médio	Desvio RCO
0	18.039.470	4.112.363	3	0,081	0,017
1	342.197	511.793	255	0,107	0,066
2	5.964.256	2.920.291	12	0,085	0,041

Fonte: O autor (2025).

Analisando as estatísticas por *Cluster* em relação ao ticket médio, a receita e a quantidade de projetos temos as seguintes percepções e resultados

Cluster 0: Projetos muito grandes:

- a) Poucos projetos (3), com ticket médio muito alto (~R\$18 milhões de reais).
- b) RCO médio baixo (8,1%), mas estável com baixa variação.
- c) 100% dos projetos com rentabilidade Alta (será apresentado em tabela a seguir).
- d) Perfil: Projetos grandes, estáveis e rentáveis. Possivelmente institucionais, estratégicos ou com recursos garantidos.
- e) Estratégias para a fundação: garantir manutenção, continuidade e fortalecimento estratégico, através do relacionamento institucional, governança diferenciada, visibilidade e marketing institucional, inovação e expansão, gestão de riscos baixa, mas continua.

Cluster 1: Projetos pequenos:

- a) Maior número de projetos (255), com ticket médio muito baixo (~R\$342 mil reais).
- b) RCO médio um pouco maior (10,7%), mas com alta variação.
- c) Alta dispersão nos resultados, projetos pequenos variam muito em desempenho.
- d) Perfil: Alta quantidade, menor impacto individual, variabilidade nos resultados. Aqui estão apresentados os projetos de maior risco e/ou oportunidade para melhoria nos processos operacionais e administrativos.

- e) Estratégias para a fundação: ganhar escala, reduzir dispersão e aumentar a eficiência desse tipo de projetos, através da padronização e automação de processos, capacitação e suporte dos coordenadores, agrupamento ou portfólio de projetos, gestão de risco e fomento interno.

Cluster 2: Projetos Médios:

- a) 12 projetos, com ticket médio (~R\$ 6 milhões de reais).
 b) RCO médio de 8,5% com variação moderada.
 c) Menor quantidade, mas importante do ponto de vista financeiro.
 d) Perfil: Projetos relevantes em valor, mas com retorno similar aos grandes (*cluster 0*), porém com mais incertezas.
 e) Estratégias para a fundação: consolidar e transformar em grandes projetos ou garantir a sustentabilidade, através da gestão ativa de desempenho, investimento em estruturação, estímulo à inovação, política de escalonamento e gestão de incertezas.

No Quadro 15, abaixo será apresentado o quadro com a distribuição das Classes de Rentabilidade por *cluster*.

QUADRO 15 – DISTRIBUIÇÃO DAS CLASSES DE RENTABILIDADE X *CLUSTER*

<i>Cluster</i>	ALTA	BAIXA	MÉDIA
0	100%	0%	0%
1	45%	41%	15%
2	75%	17%	8%

Fonte: O autor (2025).

Cluster 0: Todos os projetos são altamente rentáveis, mesmo com RCO médio moderado em (8,1%). Isso sugere que o RCO sozinho não define a classe, pode haver outros critérios envolvidos que devam ser analisados como retorno absoluto ou variações de custo.

Cluster 1: É o mais heterogêneo, com quase metade dos projetos tendo rentabilidade Alta, mas também muitos com baixa. É aqui que está a maior instabilidade, tanto projetos ruins quanto bons no aspecto de rentabilidade. Ideal para focar em estratégias e melhoria de gestão.

Cluster 2: Embora menor, tem 75% dos projetos com rentabilidade Alta, melhor que o *cluster 1* e quase tão bom quanto o *cluster 0*, mas com valores mais baixos. Pode ser um *cluster* promissor para investir ou replicar modelos.

A análise dos *clusters* culmina na identificação de vetores estratégicos para a fundação como apresentado no Quadro 16 abaixo.

QUADRO 16 – PONTOS FORTES, FRACOS E OPORTUNIDADES X *CLUSTERS*

Cluster	Pontos Fortes	Pontos Fracos	Oportunidades
0	Muito rentáveis e estáveis.	Poucos casos, difícil escalar.	Replicar esse tipo de projeto.
1	Diversidade, muitos projetos.	Alta instabilidade, risco alto.	Aplicar regras de qualificação, melhorar gestão.
2	Bons resultados com valores médios.	Pouca quantidade.	Explorar como “modelo médio eficiente”.

Fonte: O autor (2025).

Após as análises conclui-se que o *cluster* 1 deve ser o foco da gestão para migrar, enquanto os *clusters* 0 e 2 representam modelos de sucesso (alto e médio valor) a serem mantidos e, se possível replicados.

Em seguida, após a segmentação do portfólio por meio da análise de *cluster*, uma abordagem complementar de mineração de dados foi empregada para investigar padrões de concorrência entre as variáveis.

Aplicou-se o algoritmo *Apriori* com o objetivo de identificar regras de associação relevantes, as quais poderiam relacionar o tipo de projetos (ex: CURSO, TRIPARTITE etc.) e as classes de rentabilidade (ALTA, BAIXA, MÉDIA) com resultados específicos (ex: RCO Baixo ou Receita Baixa etc.).

A execução do algoritmo *Apriori* resultou em 36 regras de associação, as métricas de suporte, confiança e *lift* evidenciaram a robustez estatística das regras com um alto valor de Confiança (próximas a 1,0) indica que, se o antecedente for verdadeiro, o consequente também será verdadeiro.

Em resumo o algoritmo *Apriori* confirmou as relações já esperadas entre as variáveis de rentabilidade e as dimensões de Receita/RCO. Embora o resultado seja metodologicamente válido, ele não produziu regras de intervenção acionáveis (ex: Se o projeto for do tipo X e tiver Y, então a rentabilidade será muito alta), limitando sua contribuição direta ao apoio à tomada de decisão neste contexto.

Portanto, com a utilização do algoritmo *Apriori*, as relações encontradas, no entanto, não trouxeram percepções e padrões interessantes que possam ser utilizados para apoio à tomada de decisão.

5.5 AVALIAÇÃO

Ao analisar a base de dados da fundação e após a execução de diversas metodologias desde a análise exploratória de dados e a mineração de dados, é possível encontrar algumas informações importantes para a tomada de decisão estratégica da fundação.

Segundo Ferreira *et al.* (2021), a análise exploratória de dados segue o seguinte fluxo: Problema – Dados – Análise – Modelo, esse fluxo foi utilizado nesse projeto de pesquisa e aplicado com a busca de resolver o problema que é a sustentabilidade financeira da fundação dentro dos dados disponíveis no banco de dados do sistema de gestão da fundação.

A partir da análise dos dados, foi desenvolvido um modelo matemático com o objetivo de compreender a relação entre o RCO CALCULADO e o RCO EFETIVO.

A aplicação desse modelo permitiu identificar padrões e discrepâncias relevantes, contribuindo para uma melhor compreensão dos fatores que influenciam a remuneração efetiva da fundação.

Os resultados indicam que há situações e restrições impostas às fundações de apoio que impactam diretamente a sua capacidade de remuneração. Tais, limitações, muitas vezes de natureza normativa ou legal, comprometem a sustentabilidade financeira das fundações, mesmo quando há eficiência na execução dos projetos e esforços na captação de recursos.

Essas restrições começam a partir do momento em que a fundação não pode cobrar o preço pela prestação dos serviços prestados na gestão administrativa e financeira dos projetos, e que permitem somente a recuperação das suas despesas operacionais administrativas.

Além desse ponto, existe ainda a determinação legais e normativas e a previsão legal dos editais de fomento e pesquisa que atribuem percentuais específicos que em muitas vezes não cobrem os custos para gestão dos projetos.

Então essa é uma questão legal que até o momento não foi revista e que acaba por afetar diretamente a sustentabilidade da fundação estudada e de todas as fundações do Brasil. A única fonte de remuneração da fundação de apoio é aquela recebida pela prestação dos seus serviços de gestão administrativa e financeira de projetos em apoio às IFES ou ICTs.

Com isso a aplicação dos algoritmos de mineração de dados, incluindo técnicas de classificação, regressão, clusterização e associação, possibilitou a identificação de

padrões e situações relevantes, capazes de fornecer informações úteis ao processo de tomada de decisão estratégica por parte da fundação.

Esses métodos contribuíram para uma análise mais aprofundada dos dados operacionais, permitindo a visualização de aspectos que afetam diretamente a rentabilidade e sustentabilidade dos projetos executados.

A análise dos *clusters* e das classes de rentabilidade evidenciou que a recuperação dos custos operacionais representa um fator limitante para os resultados financeiros das fundações. Verificou-se ainda a existência de projetos com alta movimentação de recursos, mas com baixo índice de recuperação de custos operacionais.

Em contrapartida, observou-se também projetos de menor escala, mas com índices mais elevados de recuperação, ainda que estes projetos se apresentem mais suscetíveis às oscilações e flutuações do ambiente externo.

Sob à ótica legal, a sustentabilidade financeira das fundações é impactada por restrições impostas pela legislação vigente, que determina o ressarcimento apenas dos custos efetivamente incorridos na execução dos projetos.

Ao exigir das fundações uma metodologia formal de cálculo, o legislador busca garantir a comprovação dos custos administrativos e operacionais vinculados a cada projeto específico.

Essa exigência, embora promova maior transparência, acaba por limitar a capacidade de remuneração plena das instituições. Se todos os projetos ressarcissem integralmente os custos efetivos das fundações, os riscos relacionados a sustentabilidade financeira seriam significativamente reduzidos.

No entanto, o que se observa na prática são entraves legais que restringem a cobrança, seja por limites impostos em lei, por diretrizes da instituição apoiada ou por exigências dos financiadores.

Diante desse cenário, as fundações têm buscado dentro das possibilidades existentes, identificar margens de ação para promover mudanças viáveis e sustentáveis.

Uma das estratégias adotadas envolve a articulação com entidades representativas, como o CONFIES, no intuito de ampliar os percentuais de remuneração previstos por algumas agências de fomento, que atualmente não ultrapassam 5%.

Destaca-se a importância da diversificação das fontes de receita por meio da prestação de serviços, respeitando o escopo de atuação e expertise da fundação enquanto instituição de direito privado para buscar fomentar e complementar a sua captação de recursos.

O aprimoramento da eficiência interna, com foco na revisão dos fluxos e processos, bem com o credenciamento junto a outras Instituições Federais de Ensino Superior (IFES), ou Institutos de Ciência e Tecnologia (ICTs) são caminhos estratégicos que podem fortalecer a posição institucional e ampliar as possibilidades de captação de recursos.

5.6 IMPLANTAÇÃO

Com base nos resultados alcançados por meio da análise exploratória e da aplicação de técnicas de mineração de dados à base histórica de projetos geridos pela fundação, foi possível compreender dinâmicas relevantes relacionadas ao comportamento dos recursos financeiros movimentados.

Os modelos e agrupamentos gerados permitiram a identificação de padrões entre os projetos executados, evidenciando aqueles com maior rentabilidade, maior realização de receita, bem como a predominância de projetos de menor porte. Apesar disso, tais projetos apresentam oportunidades significativas de aprimoramento na gestão.

Observou-se, contudo, que grande parte das dificuldades relacionadas à recuperação dos custos operacionais está diretamente associada ao percentual máximo permitido para a remuneração da Despesa Operacional Administrativa (DOA) da fundação.

Essa restrição reflete a influência de fatores externos ao controle da instituição, tais como dispositivos legais, normativas específicas dos financiadores e critérios definidos pela administração pública ou por agências de fomento.

Diante desse cenário, cabe à fundação, em articulação com outras instituições congêneres e com o CONFIES, buscar uma atuação política que vise a revisão do entendimento por parte dos tribunais e financiadores quanto aos limites impostos à remuneração pelo ressarcimento de custos.

Apesar desse entrave, é fundamental que a fundação adote um planejamento estratégico consistente e desenvolva iniciativas voltadas ao fortalecimento de sua

sustentabilidade financeira, seja por meio do credenciamento junto a outras Instituições Federais de Ensino Superior (IFES), e Institutos de Ciência e Tecnologia (ICTs) seja pela ampliação da oferta de serviços enquanto fundação de direito privado conforme seu objeto estatutário.

Nesse contexto, torna-se essencial considerar os achados desta pesquisa, aplicando-os à gestão atual como subsídio para o aprimoramento da análise e da condução dos projetos executados, com foco na melhoria do atendimento, na melhoria dos processos e fluxos internos e na efetividade das ações institucionais.

6 RESULTADOS E ANÁLISES

A presente análise foi conduzida a partir de dados extraídos do sistema de gestão da fundação, o ERP Protheus (TOTVS), contemplando os módulos de gestão de projetos, contratos, financeiro e contabilidade.

Os dados foram importados via comando *SQL*, originando uma base histórica abrangente de todos os projetos registrados no sistema da fundação.

Durante a análise inicial da base, foram identificados diversos problemas relacionados a inconsistências, preenchimentos inadequados e ausência de informações essenciais.

Observou-se, ainda, a ausência de um padrão uniforme de lançamento por parte dos operadores do sistema, bem como a inexistência de validações obrigatórias nos campos, o que permite a inserção de dados sem qualquer padronização.

Na fase de implantação do sistema atual, embora tenham sido realizados mapeamentos prévios, optou-se por manter a configuração padrão de fábrica, que atende majoritariamente às demandas do setor privado. No entanto, dada a diversidade e complexidade dos projetos geridos por fundações de apoio, algumas customizações seriam necessárias.

Em função de restrições orçamentárias, a fundação aproveitou parte dos dados de sistemas legados, importando-os para o novo sistema com pouca ou quase nenhuma revisão, o que comprometeu a qualidade da base de dados.

Esse cenário foi constatado na base utilizada para esta pesquisa, uma vez que, apesar do recorte temporal de 2019 a 2024 representar o período com maior maturidade operacional segundo a própria fundação.

Ainda assim foram observadas falhas significativas, como campos incompletos, registros duplicados e ausência de padronização. Essas limitações exigiram um extenso processo de limpeza, normalização, validação e correção dos dados antes do início efetivo das análises.

6.1 ANÁLISE EXPLORATÓRIA DA BASE DE PROJETOS

Uma análise inicial foi realizada sobre um conjunto de 593 projetos vigentes e finalizados, compreendidos entre 01/01/2019 e 31/12/2024. Os primeiros testes

demonstraram que, sem o devido tratamento dos dados não seria possível extrair padrões confiáveis.

A base contemplava variáveis como: vigência, coordenador responsável, título do projeto, unidade gestora, tipo de recurso, fonte financiadora, entre outras. Essa análise permitiu mapear as principais unidades gestoras, fontes de financiamento mais recorrentes e os coordenadores mais produtivos.

Essas informações são úteis para o planejamento institucional, mas não são, por si só, suficientes para otimizar a captação de recursos.

Segundo Sicsú, Sarmatini e Barth (2023), a análise exploratória de dados permite identificar padrões ocultos, *outliers* e correlações, garantindo uma base sólida para a construção de modelos e a proposição de melhorias.

Considerando que o objetivo central da pesquisa é o de analisar de que forma a aplicação da análise exploratória e da mineração de dados pode contribuir para a identificação de padrões, tendências e oportunidades de melhoria nos processos operacionais e de gestão de uma fundação de apoio.

Foram aprofundadas análises quantitativas como foco na variável Recuperação de Custos Operacionais (RCO), também denominada pela lei como Despesas Operacionais Administrativas (DOA).

6.2 APLICAÇÃO DA METODOLOGIA *CRISP-DM* PROJETOS

A metodologia *CRISP-DM* foi empregada como estrutura metodológica para a condução das análises, orientou a investigação dos fatores que afetam a sustentabilidade financeira da fundação, com ênfase no desempenho dos projetos quanto ao RCO.

6.3 RESULTADO DAS ANÁLISES

De acordo com Fayyad, Piatetsky-Shapiro e Smyth (1996), a análise exploratória de dados (AED) desempenha um papel essencial no processo de descoberta de conhecimento, ao permitir que gestores extraíam informações estratégicas capazes de otimizar processos e apoiar decisões. Essa premissa se confirma na presente pesquisa, cujos resultados revelam padrões relevantes a partir da análise dos 270 projetos finalizados no período de 2019 a 2024.

Constatou-se que apenas 8,88% apresentaram RCO EFETIVO igual ou superior ao RCO CALCULADO na formalização do projeto. Os projetos do tipo SICONV destacaram-se nesse aspecto, embora exijam elevado grau de retrabalho, devido à duplicação de registros nos sistemas internos e federais.

Outros 12,60% dos projetos apresentaram RCO EFETIVO superior a 70% do RCO CALCULADO, com destaque para os projetos dos tipos NUCLEO DE CONCURSOS e TRIPARTITE.

Ainda assim, aproximadamente 78% dos projetos analisados apresentaram desempenho inferior a 70% sendo as restrições legais o principal fator limitante, como editais e normas que estabelecem tetos de remuneração entre 2,5% e 15%, frequentemente inferiores aos custos operacionais reais da fundação.

Na comparação entre RECEITA PREVISTA e RECEITA EFETIVA observou-se que 69% dos projetos executaram 75% ou mais dos recursos previstos, indicando boa capacidade de execução orçamentária.

Contudo, esse bom desempenho não se refletiu proporcionalmente na recuperação de custos, o que reforça a necessidade de revisão dos critérios de remuneração institucional.

6.4 RENTABILIDADE POR TIPO DE PROJETO

Foi calculada a rentabilidade bruta com base na razão entre RCO EFETIVO e Receita Efetiva. Projetos do tipo SICONV (2, 4 e 5), CAPTACAO2 e EVENTOS3 apresentaram os melhores índices de rentabilidade. Embora alguns desses tipos envolvam maior esforço operacional, o retorno financeiro também é maior.

6.5 APLICAÇÃO DOS ALGORITMOS DE MINERAÇÃO DE DADOS

Segundo Castro e Ferrari (2016, p. 25), “o processo de mineração corresponde à extração de minerais valiosos, como ouro e pedras preciosas de uma mina”.

De maneira análoga, os algoritmos de mineração de dados aplicados nesta pesquisa permitiram extrair informações estratégicas valiosas da base de dados da fundação, revelando padrões que contribuem para a melhoria da gestão e da tomada de decisão.

A classificação da variável Classe de Rentabilidade (ALTA, MÉDIA e BAIXA) indicou bom desempenho do modelo na classe ALTA, com Taxa de Verdadeiros Positivos ($TP\ Rate = 84,1\%$).

A classe MÉDIA apresentou maior ambiguidade, sugerindo que projetos altamente rentáveis têm características mais bem definidas que os demais.

De acordo com Goldschmidt (2015), diversas tarefas de mineração de dados podem ser adaptadas ou combinadas entre si para gerar novas abordagens mais complexas e úteis no processo de descoberta do conhecimento (*KDD*).

Neste sentido, a *clusterização* utilizando o algoritmo *k-means*, apresentou a segmentação de projetos em três *clusters* e revelou perfis distintos:

- a) *Cluster 0*: Projetos grandes, estáveis e altamente rentáveis.
- b) *Cluster 1*: Muitos projetos pequenos, com alta variabilidade e risco.
- c) *Cluster 2*: Projetos médios, com bom desempenho e potencial de replicação

Segundo Han e Kamber (2012) e Weiss e Indurkha (1998), a composição dessas tarefas é uma prática consolidada na mineração de dados, pois permite a construção de análises mais robustas e direcionadas.

A análise dos *clusters*, neste caso, possibilita o desenvolvimento de estratégias específicas de gestão para cada grupo de projetos, otimizando o acompanhamento e a alocação de recursos.

O modelo preditivo de regressão aplicado à variável RCO EFETIVO apresentou um erro médio de 6,5%, considerado aceitável para cenários operacionais e financeiros. O modelo pode ser utilizado futuramente em simulações e apoio a tomada de decisão.

A utilização do algoritmo *Apriori*, as relações encontradas, no entanto, não trouxeram percepções e padrões interessantes que possam ser utilizados para apoio à tomada de decisão.

6.6 PROPOSTA DE *FRAMEWORK* PARA GOVERNANÇA DE DADOS E APOIO À GESTÃO DE PROJETOS

Com base nos achados da presente pesquisa, propõe-se um *framework* integrado de apoio à governança de dados e à gestão estratégica de projetos em fundações de apoio vinculadas às IFES.

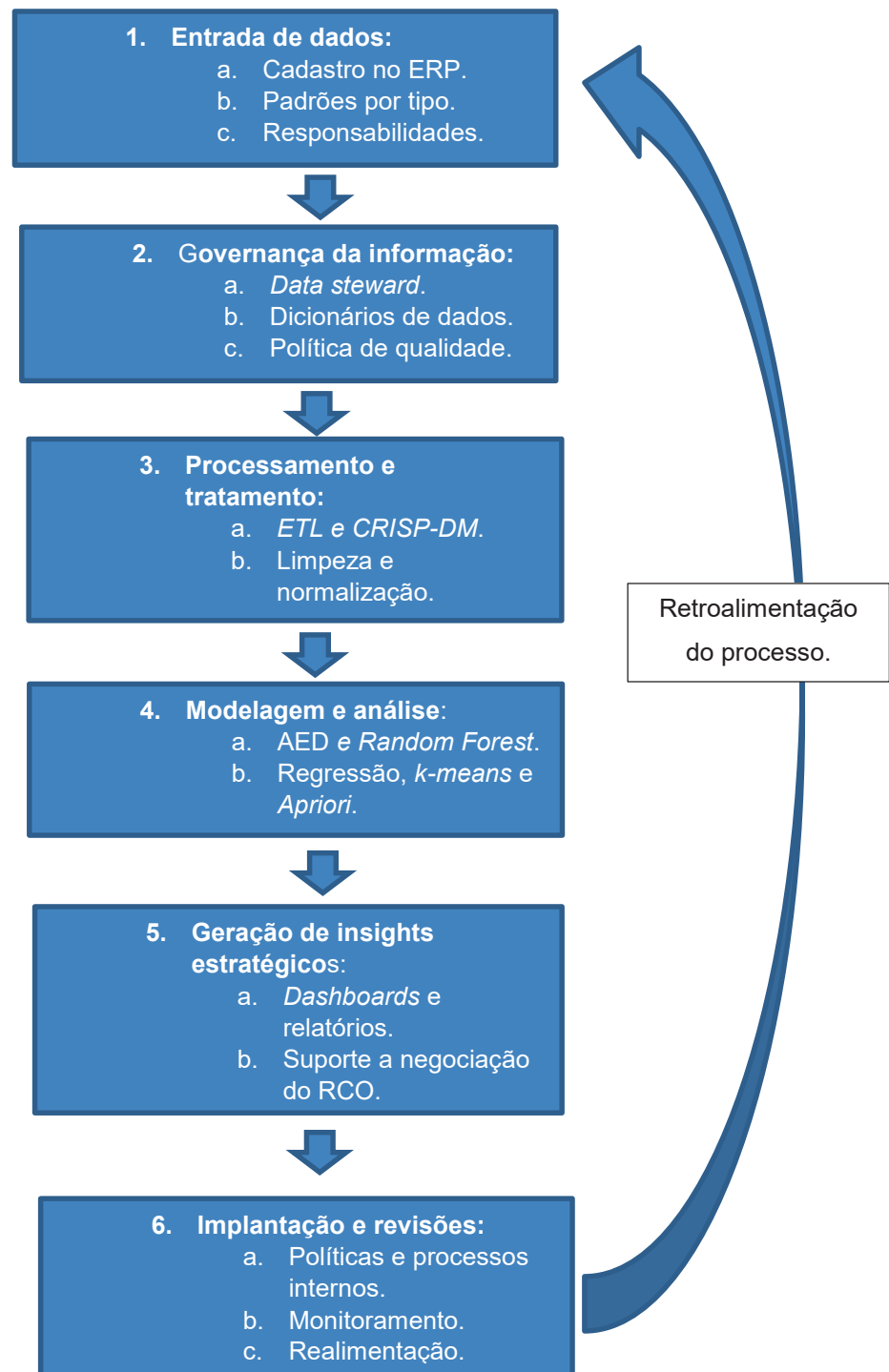
Esse modelo tem por objetivo estruturar o uso sistemático de dados operacionais para apoiar a captação de recursos, a formalização, a execução e o monitoramento de projetos, bem como promover maior controle institucional e sustentabilidade financeira.

O *Framework* foi construído a partir da aplicação da metodologia *CRISP-DM*, conforme propõe (Chapman *et al.*, 2000) e dos resultados obtidos com as técnicas de análise exploratória de dados (AED) e mineração de dados.

De acordo com Barata e Prado (2015), as organizações têm buscado, de forma crescente, garantir a integridade e a qualidade dos dados com suporte à tomada de decisões estratégicas e à geração de valor.

Nesse contexto, a Governança de Dados surge com um conjunto de processos e práticas voltadas ao gerenciamento e à manutenção das informações institucionais.

A Figura 7 o *Framework* proposto para apoio à gestão de projetos em fundações de apoio.

FIGURA 7 – *FRAMEWORK* PROPOSTO PARA APOIO À GESTÃO DE PROJETOS

Fonte: O autor (2025)

A proposta do *framework* tem como objetivo apoiar a gestão da Fundação de Apoio na adoção de um processo sistemático para a utilização dos dados institucionais em prol da tomada de decisão estratégica.

O modelo sugere um roteiro estruturado de etapas analíticas derivadas desta pesquisa, que orientam o tratamento e organização das informações operacionais até a geração de indicadores e recomendações em nível estratégico, especialmente voltadas à sustentabilidade financeira e à captação de recursos.

No âmbito da governança da informação, o *framework* integra componentes essenciais como o dicionário de dados e a política de qualidade dos dados, que atuam de forma complementar.

O dicionário de dados, de caráter operacional, descreve de maneira padronizada o significado, a origem e a estrutura das variáveis utilizadas, assegurando a consistência e a rastreabilidade das informações entre os sistemas institucionais.

Por outro lado, a política de qualidade, de natureza estratégica, define diretrizes e responsabilidades para manter a integridade, a atualidade e a confiabilidade das informações.

Adicionalmente, a aplicação de técnicas como *Random Forest*, *k-means*, Regressão Linear e *Apriori* ilustra o potencial analítico do *framework*, mas não limita suas possibilidades.

A escolha dos métodos e técnicas a serem utilizadas deve considerar as características dos dados após as etapas de processamento e tratamento, podendo outros algoritmos revelar-se mais adequados, conforme a evolução e a maturidade analítica de cada instituição.

Dessa forma, o *framework* proposto apresenta-se com um instrumento flexível e adaptável, capaz de sustentar a transformação de dados operacionais em conhecimento estratégico para aprimorar a gestão institucional.

6.6.1 Passo a passo para utilização do Framework

1. Entrada de Dados (Cadastro Inicial no ERP Protheus):

- Padrões definidos por tipo de projeto;
- Validação de campos obrigatórios;
- Responsabilidade atribuída por unidades (quem será o responsável)

2. Governança da Informação:

- Definição do responsável por dado (*data steward*);
- Criação de dicionário de dados e políticas de qualidade;

3. Processamento e Tratamento:

- Extração e transformação periódica dos dados (*ETL*);
- Aplicação da metodologia *CRISP-DM*;
- Correção de duplicidades, nulos e inconsistências.

4. Modelagem e análise:

- Aplicação de técnicas de AED (gráficos, estatísticas, rankings)
- Mineração de dados: classificação (*Random Forest*), regressão, clusterização (*k-means*), *Apriori*;
- Identificação de padrões de rentabilidade e desempenho.

5. Geração de percepções estratégicas:

- Painéis e relatórios de acompanhamento para a diretoria;
- Identificação de unidades e projetos com alto potencial de retorno;
- Apoio à negociação Institucional sobre o percentual de DOA (RCO).

6. Implantação de ações e revisões contínuas:

- Atualização das políticas e processos internos com base nas percepções;
- Monitoramento periódico da base de dados;
- Realimentação do sistema com novos indicadores.

Este *framework* poderá ser adaptado e replicado por outras fundações de apoio, respeitando as especificidades de seus sistemas de gestão e modelos de remuneração.

Seu principal diferencial está na combinação entre práticas consolidadas de mineração de dados e diretrizes de governança da informação aplicadas a contextos públicos e educacionais.

A implantação do *framework* poderá ser operacionalizada em três etapas: (1) padronização e qualificação dos cadastros de projetos; (2) designação de uma *data steward* para validação dos dados; e (3) integração contínua com *dashboards* em *Power BI*, para análise preditiva e acompanhamento de indicadores.

As principais barreiras identificadas são de natureza cultural, técnica e de governança exigindo capacitação e padronização dos processos internos.

6.7 CONSIDERAÇÕES FINAIS DA ANÁLISE

Na pesquisa verificou-se um conjunto de contribuições originais e relevantes para o campo da Ciência da Informação e para a prática institucional de fundações de apoio, tanto do ponto de vista científico quanto aplicado, alinhando-se diretamente à linha de pesquisa Informação e Gestão Organizacional do Programa de Pós-Graduação em Gestão da Informação da UFPR.

Essa linha de pesquisa prioriza investigações voltadas à produção, gestão, uso e governança da informação com recurso estratégico em ambientes institucionais e organizacionais.

6.7.1 Contribuições Científicas e Acadêmicas

1. Aplicação inovadora da metodologia *CRISP-DM* em contexto de fundação de apoio:

Esta dissertação apresenta uma aplicação inédita da metodologia *CRISP-DM* em fundações de apoio vinculadas a instituições federais de ensino, contexto ainda não explorado na literatura nacional.

A estruturação das etapas analíticas mostrou-se eficaz para extrair valor de dados operacionais, abrindo caminho para replicações em outras entidades semelhantes.

2. Proposição de *framework* de governança de dados para fundações de apoio:

Como contribuição original, a pesquisa propõe um *framework* prático adequado à realidade das fundações de apoio, fundamentado em princípios de governança da informação.

Este modelo avança sobre abordagens genéricas já consolidadas, ao oferecer diretrizes específicas para institucionalização de papéis, processos e padrões de qualidade de dados em organizações ligadas a pesquisa e inovação.

3. Integração de técnicas de AED e mineração de dados como suporte à tomada de decisão:

A dissertação demonstra, de forma pioneira, a efetividade entre Análise Exploratória e algoritmos de mineração de dados para identificar padrões de rentabilidade, desempenho financeiro e conformidade.

A integração representa uma abordagem inovadora de apoio à gestão pública e institucional, ainda pouco documentada em estudos sobre o terceiro setor e fundações de apoio.

4. Contribuição à literatura nacional sobre fundações de apoio:

O estudo preenche uma lacuna relevante na literatura brasileira ao aplicar técnicas avançadas de análise de dados à realidade de uma fundação de apoio vinculada a IFES.

Ao fazer isso, introduz uma nova perspectiva metodológica na Ciência da Informação, com potencial para inspirar futuras pesquisas e aplicações práticas.

6.7.2 Alinhamento com a Linha de Pesquisa

Gestão estratégica da informação: Ao propor mecanismos baseados em dados para melhoria dos processos de captação e controle institucional.

Governança da Informação: Por meio da criação de um modelo com princípios de padronização, definição de papéis e responsabilização pelo ciclo de vida dos dados.

Tomada de decisão: A pesquisa produziu informações a partir de dados operacionais, apoiando decisões em nível tático e estratégico.

Apoio à gestão organizacional: A dissertação contribui com métodos e ferramentas que fortalecem a capacidade de gestão em uma fundação de apoio, com potencial de replicação em outras entidades públicas e privadas.

Desta forma, os resultados obtidos fortalecem o entendimento de que a informação, quando adequadamente tratada e explorada, pode ser um diferencial para sustentabilidade, eficiência e transparência organizacional.

Os resultados obtidos reforçam que os principais entraves à sustentabilidade financeira da fundação decorrem de limitações normativas relacionadas à cobrança da DOA.

Ainda que diversos projetos demonstrem alta capacidade de execução orçamentária, o retorno financeiro à instituição é significativamente comprometido pelos percentuais máximos permitidos para ressarcimento, definidos por editais, normas institucionais e marcos legais, além das regras que muitos financiadores impõem para a comprovação para cobrança da DOA.

A legislação vigente, como a Lei nº 8.958/1994 e seu regulamento, o Decreto nº 7.423/2010, estabelece que a remuneração das fundações pode ocorrer de forma proporcional e justificada, devendo os seus custos operacionais estar claramente vinculados às atividades geridas.

Neste contexto, a informação, tratada como ativo estratégico, passa a desempenhar um papel central para a tomada de decisão gerencial. Conforme destacam Goldschmidt (2015) e Huang e Yue (2024), o uso estruturado de dados operacionais e analíticos contribui para uma gestão mais proativa, responsiva e orientada por evidências.

A análise de padrões, tendências e gargalos operacionais, como a que foi realizada neste estudo, é essencial para fundamentar decisões mais eficazes no setor público e nas organizações de apoio à ciência e tecnologia.

Adicionalmente, de acordo com Choo (2003), citado por Drummond (2008), a gestão do conhecimento organizacional atua em três arenas interdependentes: a construção de sentido, a criação do conhecimento e a tomada de decisão.

A análise de dados da fundação permitiu justamente transitar entre essas arenas, ao transformar dados brutos em informações relevantes e, posteriormente, em recomendações práticas para as áreas operacionais e estratégicas.

Com base nas análises realizadas, a fundação poderá desenvolver iniciativas voltadas a:

- a) Identificar e replicar os projetos com maior rentabilidade;
- b) Propor melhorias nas estratégias de negociação da DOA;
- c) Apoiar pleitos institucionais junto ao CONFIES, por revisão dos limites legais de ressarcimento;
- d) Promover ações internas de captação ativa com foco em projetos com maior retorno;
- e) Reavaliar fluxos e processos com vistas à redução de custos operacionais;
- f) Utilizar sua expertise em gestão de projetos para buscar credenciamento junto a outras IFES e ICTs, ampliando sua atuação institucional.

Dessa forma, a pesquisa contribui com um instrumental técnico metodológico para a gestão da informação, fortalecendo a capacidade da fundação de tomar decisões fundamentadas e orientar sua estratégia de forma mais eficaz e sustentável.

A aplicação da *CRISP-DM*, aliada à análise exploratória e técnicas de mineração de dados, demonstrou-se eficaz na identificação de padrões operacionais e financeiros.

Os resultados contribuem para o fortalecimento da governança institucional e para a definição de estratégias de captação e gestão mais sustentáveis.

7 CONSIDERAÇÕES FINAIS

Diante dos resultados obtidos, considera-se que a pesquisa respondeu ao problema proposto ao demonstrar de que forma a aplicação estruturada de técnicas de análise exploratória e mineração de dados podem apoiar o processo de tomada de decisão nas fundações de apoio?

Ao demonstrar que é possível obter informações relevantes e aplicáveis a partir desses registros, desde que a fundação adote práticas sistemáticas de análise, fundamentadas em modelos consolidados como o *CRISP-DM* e sustentadas por dados de qualidade.

A dissertação demonstrou que a aplicação de técnicas de Análise Exploratória de Dados (AED) e Mineração de Dados em registros operacionais de uma fundação de apoio vinculada a uma Instituição Federal de Ensino Superior (IFES) possibilita a extração de múltiplos conhecimentos estratégicos, com impacto direto na gestão institucional e na tomada de decisão.

Dentro os conhecimentos extraídos destacam-se:

O baixo desempenho na recuperação de custos administrativos operacionais (RCO), em apenas 8,88% dos projetos analisados apresentaram RCO igual ou superior ao volume estimado, enquanto cerca de 78% ficaram abaixo de 70%, evidenciando um padrão recorrente de insuficiência de cobertura das despesas administrativas.

A identificação de perfis de projetos por meio de *clusterização*, foram apresentados três *clusters* distintos foram formados, refletindo grupos de projetos com diferentes combinações de rentabilidade, risco e valor de repasse. Essa segmentação possibilita a fundação desenvolver estratégias diferenciadas para cada perfil, otimizando a alocação de recursos e o acompanhamento gerencial.

A necessidade de padronização e governança dos dados, os dados analisados apresentaram inconsistência de preenchimento e ausência de normalização, o que impacta negativamente a qualidade da análise.

A pesquisa propôs, como resposta prática, a criação da função de *data steward*, a padronização do cadastro de projetos e o uso sistemático de painéis analíticos interativos como instrumentos de apoio à gestão estratégica.

Esses conhecimentos foram sistematizados em um *framework* de apoio à decisão institucional, fundamentada na cultura de dados, na governança da informação e na valorização da análise preditiva como instrumento do planejamento.

Assim, a pesquisa responde pontualmente à questão proposta, ao demonstrar que a fundação de apoio pode, sim, extrair conhecimento útil de seus dados operacionais – desde que adote práticas sistemáticas de análise, estruturadas em modelos consolidados como o *CRISP-DM* e com base em dados de qualidade.

Embora é importante reconhecer as limitações do estudo. A análise concentrou-se em apenas uma fundação de apoio, o que restringe a abrangência do estudo.

O horizonte temporal delimitado (2019-2024) limita a observação de tendência de longo prazo, e a ausência de determinadas variáveis, por tratar-se a base de dados somente de projetos já encerrados, o que impediu análises mais abrangentes.

Ainda assim, os métodos aplicados demonstram elevado potencial de aplicabilidade em outras fundações também adotam o modelo de ressarcimento por Despesas Operacionais Administrativas (DOA),

O *framework* desenvolvido pode ser testado em instituições com diferentes contextos regionais e institucionais. Tal replicação permitirá avaliar a validade externa dos resultados e contribuir para a consolidação de uma prática sistemática de análise de dados no âmbito das fundações de apoio, fortalecendo a tomada de decisão estratégica no ensino superior e na pesquisa científica no Brasil.

7.1 ATENDIMENTO AOS OBJETIVOS

O objetivo geral da pesquisa consistiu em analisar o potencial da aplicação de técnicas de análise exploratória e mineração de dados na identificação de padrões, tendências e oportunidades de melhoria nos processos operacionais e de gestão de uma fundação de apoio que subsidiem ações gerenciais e processos decisórios.

Este objetivo foi plenamente atendido por meio da aplicação da metodologia *CRISP-DM*, estruturando a análise em seis etapas e utilizando técnicas de Análise Exploratória de Dados (AED) e de Mineração de Dados, que permitiram mapear variáveis-chave, visualizar padrões ocultos e gerar percepções estratégicas aplicáveis à gestão da fundação.

7.1.1 Objetivos específicos

Com vistas a evidenciar o alcance dos propósitos delineados nesta pesquisa, procede-se à análise do cumprimento dos objetivos específicos estabelecidos.

Analisar os dados disponíveis na base de dados da fundação para identificar variáveis-chave que influenciam na gestão e operação de projetos: foi realizada a análise detalhada da base de dados, identificando variáveis como tipo de recurso, tipo de vigência, rentabilidade, volume de repasse, execução financeira, RCO CALCULADO, RCO NEGOCIADO e RCO EFETIVO. Além da análise dos recursos executados nos projetos e a identificação dos projetos mais rentáveis. Esses elementos foram reconhecidos como determinantes para o desempenho e a sustentabilidade dos projetos.

Identificar padrões e relações nos dados que possam indicar oportunidades de melhoria na gestão institucional: a aplicação de técnicas de classificação, *clusterização*, regressão e regras de associação possibilitaram identificar padrões como a baixa recuperação de custos operacionais (RCO) na maioria dos projetos, a formação de três *clusters* distintos de perfis de projetos e associações não muito relevantes entre faixas de valor e desempenho financeiro.

Estabelecer recomendações baseadas nos dados analisados, desenvolver e apresentar *framework* de governança de dados, otimizar os processos internos e a tomada de decisão estratégica com foco na captação de recursos: com base nas análises, foram propostas ações o desenvolvimento de um *framework* de governança de dados com a padronização do cadastro de projetos, a criação da função de *data steward*, a implantação de painéis interativos e a revisão dos fluxos administrativos. Essas recomendações visam melhorar a governança dos dados, aumentar a eficiência na gestão e fortalecer a captação de recursos.

7.2 SUGESTÕES PARA TRABALHOS FUTUROS

A pesquisa apresentou uma limitação dos resultados apresentados pela análise de dados realizada na base de dados da fundação. No entanto, foi possível encontrar alguns padrões e percepções dos projetos que foram geridos pela fundação, inclusive questões relacionadas a limitações legais de percentual de recuperação de custos operacionais.

Um ponto importante que deve ser destacado é que embora a fundação não tenha controle direto sobre os critérios externos de captação dos recursos da DOA, ela detém total responsabilidade e expertise sobre a gestão administrativa e financeira dos projetos. Isso abrange desde a captação, negociação formalização, execução e prestação de contas, até o encerramento efetivo dos projetos.

Diante disso, compreende-se que grande parte dos esforços da fundação concentra-se em atividades operacionais internas, distribuídas em fluxos complexos e interdependentes entre diversas unidades: captação de oportunidades, apoio à formalização, gestão de projetos, execução financeira, gestão de recursos humanos, aquisições de bens e serviços, contabilidade, relatórios e prestação de contas.

Assim, surgem oportunidades significativas para aperfeiçoamento interno por meio da análise desses processos, com vistas à redução de custos operacionais, melhoria da eficiência e entrega de valor à IFES apoiada.

Diferente dos fatores externos que influenciam a definição da DOA, os fluxos dos processos operacionais e administrativos da fundação são passíveis de mensuração, monitoramento e melhoria contínua.

Portanto, recomenda-se em trabalhos futuros, que a fundação invista na aplicação de mineração de processos como ferramenta para:

- a) Identificar gargalos e atividades com maior tempo ou custo;
- b) Mapear desvios ou retrabalhos nos fluxos padrão;
- c) Avaliar o tempo médio entre etapas;
- d) Criar indicadores internos de desempenho e eficiência administrativa.

Mesmo diante da impossibilidade de interferir diretamente nos critérios de definição da DOA, a fundação poderá, através da aplicação de técnicas de mineração de processos, aprimorar sua performance organizacional, justificar de forma mais consistente os custos operacionais junto às instâncias de controle e fortalecer sua posição como instituição de referência na gestão de projetos de ensino, pesquisa, extensão e inovação.

Como desdobramento desta pesquisa, recomenda-se a realização de estudos futuros que explore o uso de técnicas de *process mining* para mapear e compreender os fluxos operacionais da fundação, identificando gargalos e oportunidades de otimização.

Além disso, o emprego de modelos de inteligência artificial generativa poderá ser investigado como ferramenta de apoio à elaboração automatizada de relatórios e simulação de cenários de decisão, ampliando o potencial analítico e preditivo do processo de gestão.

REFERÊNCIAS

- ABREU, B. O papel crucial das Fundações de Apoio às IFES E ICTs: potencialidades e contribuições para o avanço científico e tecnológico. **Revista Tópicos**, v. 2, n. 11, 2024. ISSN: 2965-6672.
- ABU-ALSONDOS, I. A. The impact of business intelligence system (BIS) on quality of strategic decision-making. **International Journal of Data and Network Science**, v. 7, n. 4, p. 1901–1912, 2023.
- ANDIFES. **Associação Nacional de Dirigentes de Instituições Federais de Ensino Superior**. Disponível em <https://www.andifes.org.br/#>. Acesso em 04 ago. 2023.
- ARGHIR, D.-C.; DUŞA, I.-G.; ONUŢĂ, M. Organizational development through Business Intelligence and Data Mining. **Database Systems Journal**, v. 10, n. 1, p. 82–99, 2019.
- BARATA, A. M. e PRADO, E. P. V. Governança de dados em organizações brasileiras. 2015, **Anais**. Porto Alegre: Sociedade Brasileira de Computação - SBC, 2015. Disponível em <http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2015/037.pdf>. Acesso em 12 ago. 2025.
- BUFREM, L. S.; ALVES, E. C. **A dinâmica da pesquisa em Ciência da Informação**. João Pessoa, PB: UFPB, 2020.
- BRASIL. **Decreto nº 7.423, de 31 de dezembro de 2010**. Regulamenta a Lei nº 8.958, de 20 de dezembro de 1994, que dispõe sobre as relações entre as instituições federais de ensino superior e de pesquisa científica e tecnológica e as fundações de apoio, e revoga o Decreto nº 5.205, de 14 de setembro de 2004. Disponível em https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/decreto/d7423.htm. Acesso em 10 jun. 2025.
- BRASIL. **Lei nº 8.958, de 20 de dezembro de 1994**. Dispõe sobre as relações entre as instituições federais de ensino superior e de pesquisa científica e tecnológica e as fundações de apoio e dá outras providências. Disponível em https://www.planalto.gov.br/ccivil_03/leis/L8958.htm. Acesso em 10 out. 2022.
- BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002**. Institui o Código Civil. Disponível em https://www.planalto.gov.br/ccivil_03/leis/2002/l10406compilada.htm. Acesso em 05 out. 2022.
- CAETANO, E. F. S.; CAMPOS, I. M. B. M. A autonomia das universidades federais na execução das receitas próprias. **Revista Brasileira de Educação**, 24, e240043, p. 9-19, 2019. Disponível em <https://doi.org/10.1590/S1413-24782019240043>.
- CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016.

CHOO, C. W. **A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões.** São Paulo: Senac, 2003.

CONFIES. **Fundação de apoio captam mais de 8 bi e batem novo recorde na receita para ciência nacional.** Site CONFIES, 2022. Disponível em <https://confies.org.br/institucional/fundacoes-de-apoio-captam-mais-de-r-8-bi-e-batem-novo-recorde-na-receita-para-ciencia-nacional/>. Acesso em 30 out. 2024.

DAFT, R. **Organizações: teoria e projetos.** São Paulo: Thomson, 2006.

DAVENPORT, T. H.; PRUSAK, L. **Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual.** 12. ed. Rio de Janeiro: Campus, 2003.

DI PIETRO, M. S. Z. **Direito administrativo.** 10 ed. São Paulo: Atlas, 1998.

DRUMMOND, R. C. **Gestão do conhecimento em organizações.** 1 ed. São Paulo: Saraiva, 2008.

e-MEC. **Cadastro Nacional de Cursos e Instituições de Educação Superior.** Disponível em <https://emec.mec.gov.br/>. Acesso em 02 ago. 2023.

FAYYAD, U.; PIATESKY-SHAPIO, G.; SMITH, P. **Knowledge discovery and data mining: towards a unifying framework.** In: *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, OR, USA, 1996. p. 82-88.

FERNANDES, R. C. **Privado, porém público – O Terceiro Setor na América Latina.** Rio de Janeiro: Relume, 1994.

FERREIRA, R. G. C.; MIRANDA, L. B. A.; PINTO, R. A. *et al.* **Preparação e Análise Exploratória de Dados.** Porto Alegre: SAGAH, 2021. E-book. p.15. ISBN 9786556902890. Disponível em <https://integrada.minhabiblioteca.com.br/reader/books/9786556902890/>. Acesso em 06 jan.2025.

FRANK, E.; HALL, M. A.; WITTEN, I. H., **The WEKA Workbench.** In: Online appendix for Data mining: Practical Machine Learning Tools and Techniques, 2016. Disponível em https://ml.cms.waikato.ac.nz/weka/Witten_et_al_2016_appendix.pdf. Acesso em 01 ago. 2025.

GIL, A.C. **Como Elaborar Projetos de Pesquisa.** 5. ed. São Paulo: Atlas, 2010.

GOLDSCHMIDT, R. **Data mining: conceitos, técnicas, algoritmos, orientações e aplicações.** 2 ed. Rio de Janeiro: Elsevier, 2015.

HUANG, S.; YUE, H. Optimizing Strategic Decision-Making in Multinational Corporations through Data Mining-Based User Behavior Analysis. **Journal of System and Management Sciences**, v. 14, n. 1, p. 579–591, 2024. Success Culture Press.

LAKATOS, E. M.; MARCONI, M. A. **Metodologia científica**. 6. ed. São Paulo: Atlas, 2011.

MARQUES, É. V.; CORREIA NETO, J. F. **Tomada de decisões gerenciais com analítica de dados: aplicações práticas com Excel**. Rio de Janeiro: Alta Books, 2020.

MENDONÇA, L. R.; MACHADO F.; PINHEIRO, C. A. Governança nas organizações do terceiro setor: considerações teóricas. **Revista de Administração**, USP, São Paulo, n.4, p.302-308, out./nov./dez. 2004.

MOUSSAS, K.; HAFIANE, J.; ACHABA, A. Business intelligence and its pivotal role in organizational performance: An exhaustive literature review. **Journal of Autonomous Intelligence**, v. 7, n. 4, 2024. Frontier Scientific Publishing. Disponível em <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184690970&doi=10.32629%2fjai.v7i4.1286&partnerID=40&md5=e37f72fe93fed58a0b580ef6a0be8f6f>. Acesso em 10 set. 2024.

PAES, J. E. S. **Fundações, Associações e Entidades de Interesse Social – Aspectos jurídicos, administrativos, contábeis, trabalhistas e tributários**. 9. ed. Rio de Janeiro: Forense, 2017.

PEREIRA, R. M. **Terceiro Setor – identificando-se suas organizações e propondo um novo conceito**. Disponível em <http://www.oabsp.org.br/comissoes2010/direito-terceiro-setor/artigos>. Acesso em 05 set. 2022.

PERES, S. M.; FANTINATO, M.; SANTOS, E. A. P. **Mineração de Processos: do que se trata? E o Brasil, está no jogo?** Computação Brasil, [S. l.], n. 49, p. 6–9, 2023. Disponível em <https://journals-sol.sbc.org.br/index.php/comp-br/article/view/4053>. Acesso em 15 jul. 2025.

RAMOS, J. L. C.; RODRIGUES, R. L.; SILVA, J. C. S.; OLIVEIRA, P. L. S. de. CRISP-DM: uma proposta de adaptação do Modelo CRISP-DM para mineração de dados educacionais. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 31., 2020, Online. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 1092-1101. DOI: <https://doi.org/10.5753/cbie.sbie.2020.1092>.

RAZZOLINI FILHO, E. **Gestão da informação e competências necessárias ao gestor**. São Paulo: SBGC, 2009, v. 1.

REZENDE, D. A.; ABREU, A. F. **Tecnologia da Informação Aplicada a Sistemas de Informação Empresariais**. 9. ed. São Paulo: Atlas, 2013.

SICSÚ, A. L.; SAMARTINI, A.; BARTH, N. L. **Técnicas de machine learning**. 1. Ed. São Paulo: Blücher, 2023.

SHEARER, Colin. The CRISP-DM Model: the new blueprint for data mining. **Journal of Data warehousing**, v. 5, n. 4, p. 13–22, 2000.

STANKEVIX, G. **Análise exploratória de Dados – Uma aplicação do pensamento estatístico em Python**. Medium, 10 fev. 2020. Disponível em <https://medium.com/@gabriel.stankevix/analise-explorat%C3%B3ria-de-dados-732007ddbaf>. Acesso em 05 ago. 2025.

STONER, J. A. F.; FREEMAN, R. E. **Administração**. 5 ed. Rio de Janeiro: LTC, 1999.

TAKEUCHI, H.; NONAKA, I. **Gestão do conhecimento**. Porto Alegre, Bookman, 2008.

TARAPANOFF, K. **Inteligência, informação e conhecimento em corporação**. Brasília: IBICT, UNESCO, 2006.

TEIXEIRA, M.C. FREITAS DUARTE, A.M. **Engineering information: conceptual elements related information management and information systems**. EDULEARN14 *Proceedings*, pp. 6909-6915, 2014.

UFPR. **Resolução nº 41/17-COPLAD**. Disponível em <https://soc.ufpr.br/wp-content/uploads/2017/12/Res-41-17-COPLAD-reformula%C3%A7%C3%A3o-Res.-17-11-alt.-pela-Res-30-18-COPLAD-final.pdf>. Acesso em 02 ago. 2023.

UFPR. **Universidade Federal do Paraná**. Disponível em <https://ufpr.br/missao-visao-valores/>. Acesso em 02 ago. 2023.

YU, A. S. O. **Tomada de decisão nas organizações: uma visão multidisciplinar**. São Paulo: Saraiva, 2011 – 1ª Edição.