



De 11 a 14 - NOVEMBRO DE 2024

**Inteligência Artificial
na Gestão de Operações:**
Limitações e possibilidades



MINERAÇÃO DE DADOS PARA ITENS DE TREINAMENTOS DO DEPARTAMENTO DE TI DE UMA EMPRESA DO SETOR AÉREO POR MEIO DA METODOLOGIA CRISP-DM

ANA CLAUDIA RENÓ FERREIRA LISBOA - anaclaudia.reno@gmail.com
UNIVERSIDADE FEDERAL DE ITAJUBÁ- UNIFEI

DOUGLAS OLIVEIRA PEDROZO - douglas.pedrozo@yahoo.com.br
UNIVERSIDADE ESTADUAL PAULISTA- UNESP

ALEXANDRE FERREIRA DE PINHO - pinho@unifei.edu.br
UNIVERSIDADE FEDERAL DE ITAJUBÁ- UNIFEI

FABIO FAVARETTO - fabio.favaretto@unifei.edu.br
UNIVERSIDADE FEDERAL DE ITAJUBÁ- UNIFEI

ÁREA: 3. PESQUISA OPERACIONAL
SUBÁREA: 3.3 – PROCESSOS DECISÓRIOS

RESUMO: O OBJETIVO DESTE ESTUDO É CRIAR VISIBILIDADE E INDICADORES PARA APRESENTAÇÃO DAS INFORMAÇÕES DE FORMA AMIGÁVEL, FAVORECENDO TOMADA DE DECISÕES DA LIDERANÇA DE UMA EMPRESA DO SETOR AÉREO EMPREGANDO O MODELO DE PROCESSO CRISP-DM. O ESTUDO FOI REALIZADO NO DEPARTAMENTO DE TI DA EMPRESA DO SETOR AÉREO E ENVOLVE A APLICAÇÃO DA METODOLOGIA DE PROCESSO CRISP-DM. OS DADOS FORAM OBTIDOS EM BASE DE DADOS DO SISTEMA DE LEARNING MANAGEMENT SYSTEM (LMS) DO DEPARTAMENTO DE RH DA EMPRESA. O PROCESSO FOI COMPOSTO POR SEIS FASES: 1. ENTENDIMENTO DO PROBLEMA, 2. ENTENDIMENTO DOS DADOS, 3. PREPARAÇÃO DE DADOS, 4. MODELAGEM, 5. AVALIAÇÃO E 6. IMPLANTAÇÃO. FOI CONSTRUÍDA UMA SOLUÇÃO PARA ANÁLISE RELACIONADA ÀS INFORMAÇÕES REFERENTES AOS ITENS DE CAPACITAÇÃO DO TIME DE TI DA EMPRESA, TAIS COMO CUSTO TOTAL OU POR SUPERVISÃO, CATEGORIA (TREINAMENTOS INTERNO, EXTERNO E IN COMPANY), TEMAS DOS TREINAMENTOS SEGMENTADOS PELOS PILARES DE AGILIDADE, CULTURA, PROCESSOS E TECNOLOGIA, ALÉM DA QUANTIDADE DE PROFISSIONAIS TREINADOS. OS RESULTADOS OBTIDOS PODEM SER EMPREGADOS POR OUTROS DEPARTAMENTOS DA EMPRESA, E ATÉ MESMO POR OUTRAS EMPRESAS COMO FORMA DE PASSAR CONHECIMENTO POR MEIO DE CAPACITAÇÃO DOS COLABORADORES.

PALAVRAS-CHAVES: TOMADAS DE DECISÃO, TREINAMENTO EMPRESARIAL, DESENVOLVIMENTO EMPRESARIAL.

1. INTRODUÇÃO

Processar grandes quantidades de dados para apoiar os processos de decisão está se tornando uma atenção crescente na Tecnologia de Informação corporativa. Informações utilizadas no processo de decisão, principalmente relacionado as estratégias, envolvem conhecimento que podem ser fornecidos pelo sistema de informação da organização, da qual se realizará a análise de indicadores. Através dessa análise, se obtém informações valiosas dos dados que direcionaram na tomada de decisão (Krcmar; Helmut, 2015).

A Inteligência de Negócios também se caracteriza pelo uso de um ambiente de consultas ad-hoc para um fim específico. Isso pode ser obtido diretamente através de consultas não estruturadas ou através de Mineração de Dados (Plotnikova; Dumas; Milani, 2020). As atividades de Mineração de Dados constituem um processo destinado à análise de grandes bancos de dados, com o objetivo de extrair informações e conhecimentos que podem ser precisos e potencialmente úteis para os profissionais envolvidos na tomada de decisões e na solução de problemas (Vercellis, 2011). Este ambiente é criado por meio de um processo denominado de *Data Warehouse* (DW) que utiliza um Modelo Dimensional que mostra as consultas que poderão ser feitas no banco de dados criado para esta finalidade.

O modelo dimensional (ou multidimensional) é o “projeto” do DW. Ele mostra todos os indicadores e todas as dimensões do ambiente de consultas de qualquer conjunto de dimensões com qualquer conjunto de indicadores. Uma de suas características é que todas as dimensões devem ser compatíveis com todos indicadores para atender as necessidades dos gestores e a disponibilidade de dados, que servirá de um guia para o desenvolvimento do DW. O modelo dimensional se utiliza de elementos, que podem ser de diferentes dimensões, como o tempo, local, objeto e pessoas (Saxena; Agarwal, 2014).

No entanto, seguir rigorosamente uma metodologia de projeto pode ser um desafio para o processo de Mineração de dados, tendo em vista, que os modelos de processos não estão bem estabelecidos (Schröer; Kruse; Gómez, 2021). A metodologia padrão utilizada é denominada de CRISP-DM ou *Cross-Industry Standard Process of Data Mining* (Processo Padrão Inter-Indústrias para Mineração de Dados). Esse modelo é neutro, documentado e disponível livremente para utilização (Martínez-Plumed et al., 2019). A mineração de dados é usada para analisar e melhorar os processos de produção (Schnell et al., 2019). A aplicação dessa metodologia do CRISP-DM pode contribuir para o setor de serviços na área de TI de empresas do setor aéreo, tendo em vista que estudo que abordem essa área são escassos.

O problema a ser tratado com a mineração dos dados está relacionado à falta de visibilidade da liderança em relação às informações dos treinamentos ministrados em 2021,

para tomada de decisão com maior agilidade em relação a estimativa para projetos futuros, informação de atendimento ou não à meta de aprendizagem estipulada para o ano de 2021 e gaps de aprendizagem para serem tratados em 2022.

Diante disso, este estudo tem por objetivo criar visibilidade e indicadores para apresentação das informações de forma amigável, favorecendo tomada de decisões da liderança de uma empresa do setor aéreo empregando o modelo de processo CRISP-DM.

2. REVISÃO BIBLIOGRÁFICA

A análise de indicadores é utilizada pela inteligência de negócios, que é um processo, que utiliza de técnicas, tecnologias e softwares para o seu desenvolvimento. A sua abrangência pode incluir métodos, pessoas e as informações a serem tratadas. Ela realiza a análise descritiva para formular conclusões sobre desempenho histórico e atual, proporcionando contexto em torno de mudanças nos principais indicadores de desempenho. A inteligência de negócios inclui prática de análise prescritiva e preditiva, que ajudam a orientar os tomadores de decisão sobre potenciais resultados futuros, tendo em vista que os indicadores devem ser necessários e suficientes para suportar um conjunto de decisões (Pastrana; Tobón, 2020).

A Inteligência de Negócios possui as seguintes características: capacidade de extrair e integrar dados de múltiplas fontes; a transformação dos registros em informação útil para o conhecimento empresarial; a valorização da experiência; a análise de dados contextualizados; a procura de relações de causa e efeito, trabalhando com hipóteses e desenvolvendo estratégias e ações competitivas (Negash, 2004).

No final do século passado, a aplicação sistemática de técnicas de mineração de dados para extrair o conhecimento a partir de dados estava se tornando cada vez mais comum na indústria, algumas empresas e instituições perceberam a necessidade de unir forças para identificar boas práticas, bem como erros em suas experiências passadas. Nesse momento, uma equipe de engenheiros experientes em mineração de dados desenvolveu uma metodologia de mineração de dados de aplicação geral. Em 1999, a primeira versão do CRISP-DM, foi introduzida (Chapman et al., 2000). Essa metodologia foi concebida para catalogar e orientar as etapas mais comuns em projetos de mineração de dados.

O CRISP-DM revelou-se muito útil em projetos caracterizados por aplicarem metodologias de Mineração de Dados para resolução de problemas. Ele é um modelo de processo independente, de livre acesso, com as seis fases iterativas de entendimento dos negócios para implantação, que envolvem: 1. Entendimento do problema: fase da qual ocorre

a definição do problema a ser resolvido e propor as metas desejadas; 2. Entendimento dos dados: esta etapa envolve a coleta de dados de suas fontes, explorá-los e descrevê-los e verificar a qualidade desses dados; 3 - Preparação dos dados: ocorre a seleção e extração dos dados necessários para serem posteriormente modelados; 4. Modelagem: momento em que ocorre a seleção da técnica do modelo a ser utilizado, constituindo uma fase de teste. Geralmente a escolha do modelo vai depender do problema, dos negócios e dos dados; 5. Validação: ocorre a análise explanatória das descobertas da modelagem. Essa etapa deve ser realizada pelas pessoas que fazem parte do contexto do problema, para apontar se realmente o modelo atende aos objetivos definidos; e 6. Implantação: momento da implantação em produção. Pode ser um relatório final ou um componente de software. Essa fase consiste no planejamento da implantação, monitoramento e manutenção (Chapman et al., 2000).

Schröer, Kruse e Gómez (2021) realizaram uma revisão sistemática para verificar as pesquisas que tem sido realizadas empregando o modelo de processo CRISP-DM, e selecionaram 24 estudos que foram realizados entre os anos de 2017 e 2019, envolvendo diferentes setores que tem empregado essa metodologia, como na saúde, educação, setor público, tecnologia de informação, dentre outras. Todavia, estudos envolvendo o CRISP-DM no setor de aviação são mais raros, principalmente o que envolve a visibilidade e indicadores sobre o custo e quantidade de treinamentos ministrados por empresas deste setor.

Schnell et al. (2019) afirmam que a utilização do CRSIP-DM pode otimizar custos operacionais e evitar desperdícios. Nesse sentido, esses autores avaliaram a aplicabilidade do método na produção de células de bateria de íon-lítio. Por meio do processo CRISP-DM foi deduzida a influência dos parâmetros do processo na qualidade do produto final e assim foi possível implantação de tomadas de decisão mais assertivas.

3. METODOLOGIA

3.1 Descrição do estudo

O presente estudo foi realizado no mês de maio, no Departamento de TI da empresa do setor de aviação e envolve a aplicação da metodologia de processo CRISP-DM proposta por Chapman et al. (2000) e Shearer (2000) em dados obtidos da empresa. Os dados foram obtidos através da extração das informações da base de dados do sistema *de Learning Management System* (LMS) do departamento de RH da empresa. Os envolvidos no projeto foram profissionais do RH, responsável pela plataforma LMS e um profissional de TI, responsável pelo gerenciamento do orçamento dos treinamentos do departamento.

3.2 Fases de processamento de dados

A primeira etapa consistiu em conhecer e compreender os objetivos do projeto de mineração a partir da perspectiva do negócio, sendo esse conhecimento transformado em um problema de mineração de dados. Essa etapa é considerada como uma das mais importantes do processo. Foi desenvolvido um plano preliminar do projeto de mineração para busca dos objetivos a serem alcançados com a resolução do problema.

Na segunda etapa, ocorreu a coleta e a compreensão dos dados, identificando possíveis problemas de qualidade. Esse passo é essencial para evitar problemas inesperados na etapa seguinte, que é considerada a fase mais longa de um projeto. Nesse momento, também foi realizado o alinhamento dos envolvidos no processo em relação ao conhecimento dos dados.

Na terceira etapa ocorreu a preparação dos dados que abrangeu todas as atividades necessárias para construir o conjunto de dados final, a partir dos dados brutos iniciais. Esses dados preparados foram utilizados para a modelagem na etapa seguinte. Essa preparação é susceptível de ser executada várias vezes e em ordem preestabelecida. As tarefas incluíram a seleção de tabelas, registros e atributos, assim como a transformação e a limpeza dos dados. Essa etapa é considerada a mais longa do processo, do qual pode consumir de 50 a 70% do tempo de um projeto.

Após a preparação dos dados, ocorreu a modelagem, através do qual foi avaliado e selecionado a melhor técnica de modelagem para resolução do problema a ser solucionado. Nesse processo, foram executados vários modelos, usando as configurações padrões com os ajustes dos parâmetros para valores otimizados. Nesse momento, pode haver retornos na fase anterior para manipulações de dados exigidas pelo modelo.

Antes de proceder à implantação definitiva do modelo construído, este foi avaliado mais profundamente e revisto a sua construção para ter a certeza que atinge, adequadamente, os objetivos propostos. Assim, essa avaliação, garante que a organização pode utilizar os resultados obtidos e os conhecimentos descobertos. Vale ressaltar, que neste momento houve uma avaliação para verificar se algum problema de relevância foi desconsiderado anteriormente, tendo em vista que ainda era possível retornar às etapas anteriores.

Na última etapa ocorreu a implantação do modelo, na qual os conhecimentos descobertos foram usados para proporcionar melhorias na organização. Nessa fase, todo o conhecimento adquirido foi organizado e apresentado de uma forma a possibilitar o uso efetivo dentro dos processos de tomada de decisão (Shearer, 2000; Chapman et al., 2000).

Para o estudo utilizou-se o processo de mineração de dados CRISP-DM - *Cross Industry Standard Process for Data Mining*, onde as etapas compreendem:

1. Entendimento do problema
2. Entendimento dos dados
3. Preparação de dados
4. Modelagem
5. Avaliação
6. Implantação

4. RESULTADOS E DISCUSSÃO

4.1 Entendimento do problema

Nesse artigo será abordado os dados referente aos treinamentos ministrados, no ano de 2021, para os profissionais da área de TI de uma empresa do setor aéreo. A decisão analisada está relacionada às informações referentes aos itens de capacitação desse time, bem como custo, líder, categoria (treinamentos interno, externo e *in company*), pilar (Agilidade, Cultura, Processos e Tecnologia) e quantidade de profissionais treinados.

- Origem dos dados: Ferramenta LMS - *Learning Management System*
- Mais de 5.000 registros no ano;
- O que estava disponível aos gestores: Três grandes tabelas, com campos diferentes, extraídas do banco de dados da aplicação LMS (gerada pelo profissional do RH);
- Planilhas geradas – Por serem tabelas diferentes não tinha uma visibilidade única das informações referentes aos treinamentos. A junção das informações era feita de forma manual e suscetível a erros.

Em suma, existia a carência de visibilidade (relatórios e gráficos), principalmente quanto aos treinamentos e custos correspondentes a cada supervisão (líder). Necessitava-se de uma ferramenta capaz de criar visibilidades que abrangessem os dados de forma direta e organizada, capaz de consolidar os principais indicadores.

4.2 Entendimento dos dados

Os dados originais consistiam principalmente da extração de três planilhas, geradas a partir da aplicação de LMS da empresa.

1. A primeira tabela possui onze colunas, referente à extração da base de dados dos itens de treinamentos EXTERNOS, são elas:

Identificador: Chave primária;

Título do evento: Item de treinamento;

Nome do participante: Colaborador inscrito no treinamento;

Centro de custo: Código orçamentário utilizado por cada colaborador para débito das despesas;

Gestor imediato: Líder do colaborador;

Unidade: Local de alocação do colaborador,

Investimento: Custo do item de treinamento;

Conversão para reais: Aplicável quando necessário converter dólar para real;

Custo real convertido com impostos (em R\$): Custo final acrescido dos impostos;

Valor orçado: Estimativa de custo do item de treinamento;

Carga horária: Duração do item de treinamento.

2. A segunda tabela possui nove colunas, referente à extração da base de dados dos itens de treinamentos *IN COMPANY*, são elas:

Identificador: Chave primária;

Título do evento: Item de treinamento;

Nome do participante: Colaborador inscrito no treinamento;

Centro de custo: Código orçamentário utilizado por cada colaborador para débito das despesas de acordo com seu departamento de alocação;

Investimento: Custo do item de treinamento;

Conversão para reais: Aplicável quando necessário converter dólar para real;

Custo real convertido com impostos (em R\$): Custo final acrescido dos impostos;

Valor orçado: Estimativa de custo do item de treinamento;

Carga horária: Duração do item de treinamento.

3. A terceira tabela possui nove colunas, referente à extração da base de dados dos itens de treinamentos *INTERNOS*, são elas:

Identificador: Chave primária;

Título do evento: Item de treinamento;

E-mail do participante: E-mail do colaborador inscrito no treinamento;

Vice-presidência: Código da vice-presidência em que o colaborador está alocado.

Investimento: Custo total do treinamento, não rateado por colaborador;

Conversão para reais: Aplicável quando necessário converter dólar para real;

Custo real convertido com impostos (em R\$): Custo final acrescido dos impostos;

Valor orçado: Estimativa de custo do item de treinamento;

Carga horária: Duração do item de treinamento.

Além dos dados fornecidos inicialmente, percebeu-se a necessidade de uma dimensão que relacionasse o nome do colaborador com o seu e-mail, vice-presidência, centro de custo e gestor imediato, pois nem todas as tabelas possuíam todas essas informações, então foi fornecida mais uma tabela, extraída do sistema SAP HR, contendo seis colunas:

Nome: Nome do colaborador;

E-mail: Endereço eletrônico do colaborador;

Vice-presidência: Código da vice-presidência em que o colaborador está alocado;

Unidade: Local de alocação do colaborador;

Centro de custo: Código orçamentário utilizado por cada colaborador para débito das despesas de acordo com seu departamento de alocação;

Gestor imediato: Gestor do colaborador.

4.3 Preparação dos dados

Para a consolidação das origens de dados utilizou-se duas fontes:

- Sistema LMS – extração de três planilhas (treinamentos externos, *in company* e internos)
- SAP HR – extração de uma planilha (dados do colaborador).

No primeiro momento, todas as informações e dados foram coletados e analisados. As planilhas foram disponibilizadas em um diretório compartilhado na rede da empresa. E o software escolhido para o trabalho foi o Microsoft Excel, que permite a composição dinâmica de consultas em forma de relatórios e gráficos, além da possibilidade de uso de filtros, ordenação e visualização dos indicadores através de contagens, médias e outros.

Foi feito o tratamento desses dados (limpeza e correção dos dados), exclusões de colunas que não foram utilizadas e a inclusão de colunas para a criação da *flat table*.

4.4 Modelagem

Novas colunas foram criadas para compor as informações necessárias para a tomada de decisão:

- Categoria: Referente a origem de cada planilha extraída da aplicação LMS (Externo, *In Company* e Interno);

- Pilar: Referente ao assunto de cada item de treinamento (Agilidade, Cultura, Processo, Tecnologia e Renovação (certificação)).

Também considerou-se o preenchimento das informações referente aos campos: E-mail, Vice-presidência, Centro de custo, Unidade e Gestor imediato a partir do campo Nome (dados extraídos do sistema SAP) para as extrações que não continham tais informações e, deste modo, viabilizar a composição da *flat table* (Quadro 1).

QUADRO 1 - Campos da Flat Table.

Coluna	Descrição
Identificador	Chave primária
Título do evento	Item de treinamento
Nome do participante	Colaborador inscrito no treinamento
E-mail	Endereço eletrônico do colaborador
Categoria	Conforme origem de cada planilha extraída da aplicação LMS (Externo, In Company e Interno)
Pilar	Tema de cada item de treinamento (Agilidade, Cultura, Processo, Tecnologia e Renovação)
Centro de custo	Código orçamentário utilizado por cada colaborador para débito das despesas, conforme cada supervisão
Gestor imediato	Líder do colaborador
Unidade	Qual a unidade de alocação do colaborador
Investimento	Custo do item de treinamento
Conversão para reais	Aplicável quando necessário converter dólar para real
Custo real convertido com impostos (em R\$)	Custo final acrescido dos impostos
Valor orçado	Estimativa de custo do item de treinamento
Carga horária	Duração do item de treinamento.

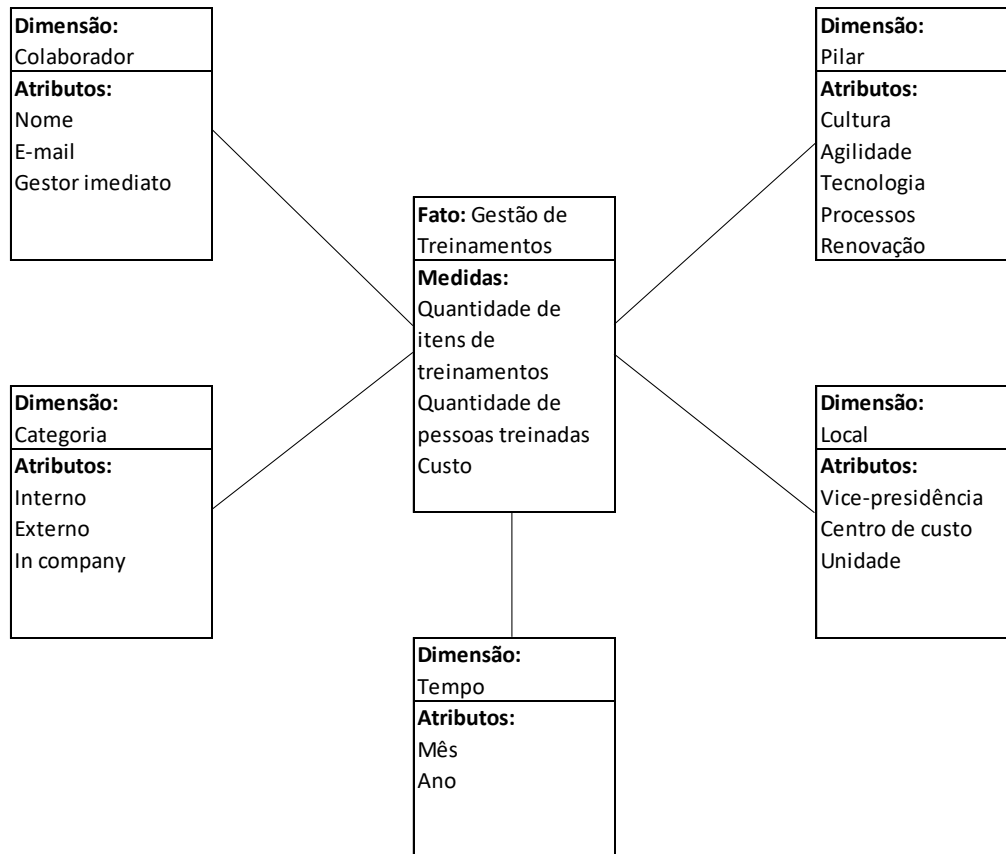
Fonte: elaborado pelos autores

O principal indicador acompanhado pela área de TI da empresa, para a qual foi desenvolvido esse modelo, foi o custo total ou por supervisão dos itens de treinamentos, bem como o total de treinamentos e participantes em cada categoria.

As análises relacionaram as medidas e as dimensões para obtenção do resultado esperado a partir da pesquisa. O *Data Wharehousing* permitiu ao usuário criar as análises de forma rápida e amigável, sem necessidade de especialista. O usuário tem a liberdade de criar

livremente suas análises conforme os objetivos de análise para mapear determinadas situações. A figura 1 apresenta o modelo dimensional.

FIGURA 1 - Modelo Dimensional.



Fonte: elaborado pelos autores

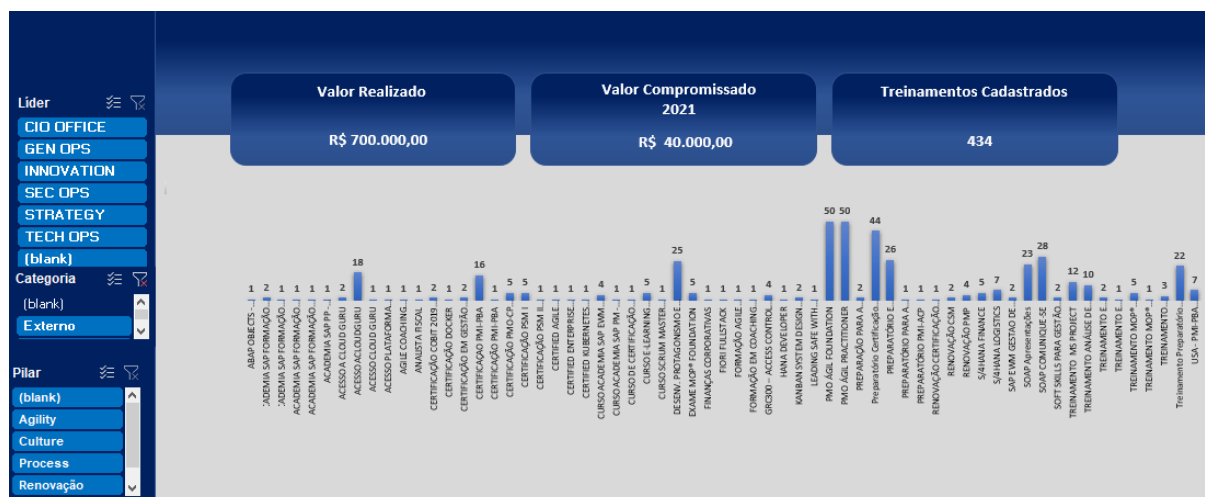
No modelo dimensional acima, verifica-se que ao centro se localiza o fator em estudo, que é a gestão de treinamentos, do qual envolve a quantidade de itens de treinamento, a quantidade de pessoas treinadas e o custo. Além disso, existem outras cinco dimensões ligadas diretamente ao fato de treinamentos: tempo, local, categoria, colaborador e pilar.

Foram gerados painéis de visibilidade de todos os treinamentos ministrados para os colaboradores de todas as supervisões, considerando as categorias:

- Externo - parceiro contratado para treinamento individual e
- *In company* - parceiro contratado para treinamento de turma fechada *on site*.

Na visibilidade também foi considerado os pilares Agilidade, Cultura, Processos, Renovação (certificações) e Tecnologia, bem como a apresentação do custo total investido (Figura 2).

FIGURA 2 - Visibilidade 1.

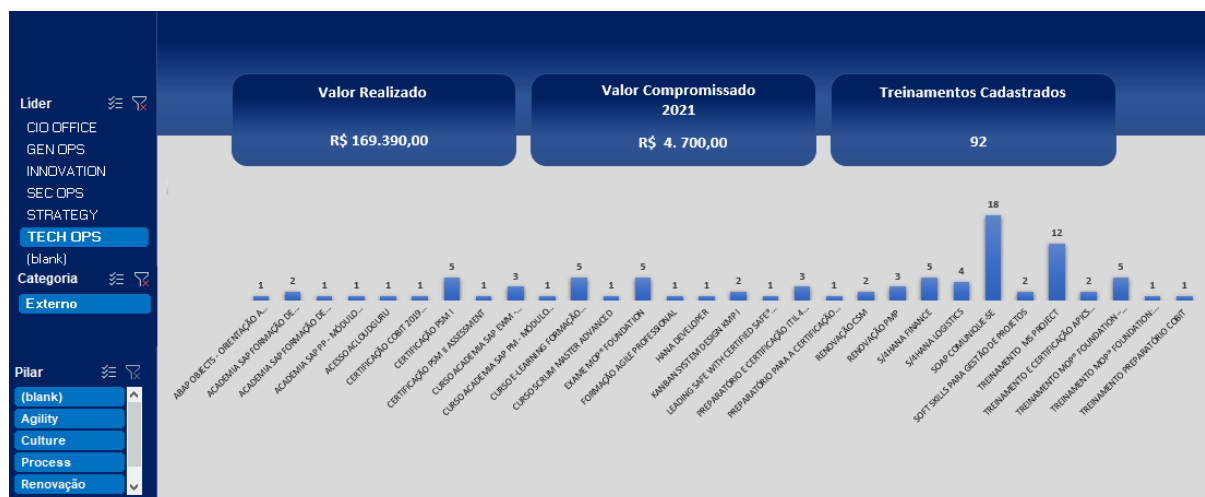


Fonte: elaborado pelos autores

Na figura 2 observa-se que o valor gasto com 434 treinamentos realizados no ano de 2021 foi de 700.000,00. Dentre os treinamentos, merece destaque os: PMO ágil *foundation* (50), PMO ágil practitioner e preparatório certificação (44). De acordo com Hamad e Fayoumi (2018), o *Project Management Office* (PMO) é a fonte de mudança para a inovação dentro de qualquer organização. Determina o sucesso das funções de gerenciamento de projetos e auxilia na obtenção dos objetivos.

O painel exposto na figura 3 foi elaborado para possibilitar a verificação dinâmica das possíveis variações de pesquisa, por exemplo, somente treinamentos da supervisão TECH OPS, nas categorias Interno e Externo, nos pilares Agilidade, Cultura e Tecnologia e, assim, apresentado o custo associado ao grupo de treinamentos a partir da seleção da pesquisa.

FIGURA 3 - Visibilidade 2.

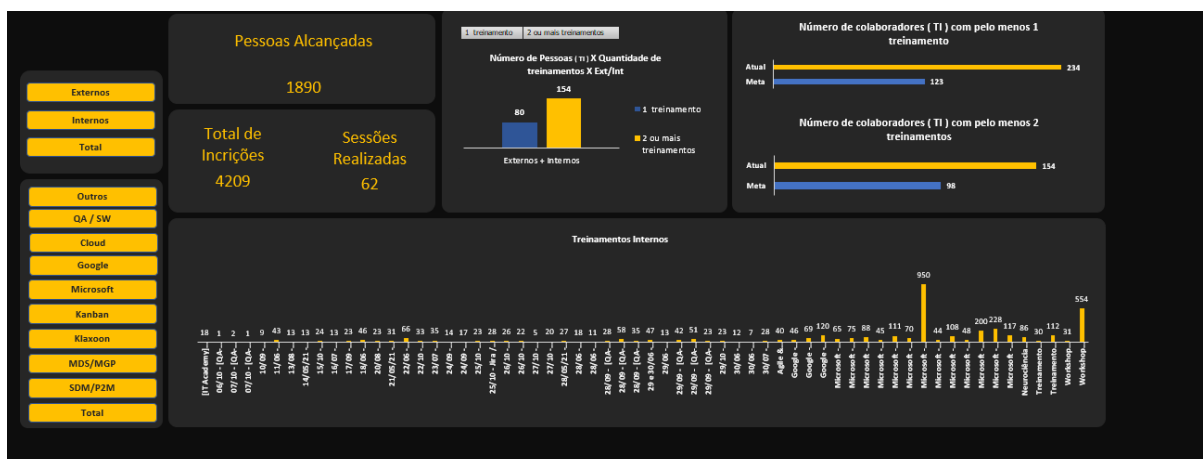


Fonte: elaborado pelos autores

TECH OPS, o termo se refere às tecnologias e operações e foi primeiro descrito por Norment, Schwenke e Kohlberg (1966). Nesse sentido, as organizações vêm passando nos últimos anos por crescentes transformações, principalmente relacionada aos usos de tecnologias, e para acompanhar esse processo, é essencial os treinamentos dos funcionários que estão operando essas novas tecnologias (Leavitt, 2013).

Na figura 4 está apresentado a quantidade de profissionais impactados (Pessoas Alcançadas: Funcionários diretos e Terceiros), total de inscritos nos treinamentos e quantas sessões realizadas, além do comparativo entre a meta estipulada pela área, referente a quantidade de pessoas treinadas, *versus* o resultado alcançado (parte superior do painel). Um total de 154 pessoas participaram de mais de um treinamento, além disso, é possível registrar que as metas de participação dos colaboradores de TI, tanto para 1 treinamento ou mais que 1, foram superadas, o que demonstra empenho dos envolvidos em querer adquirir conhecimento.

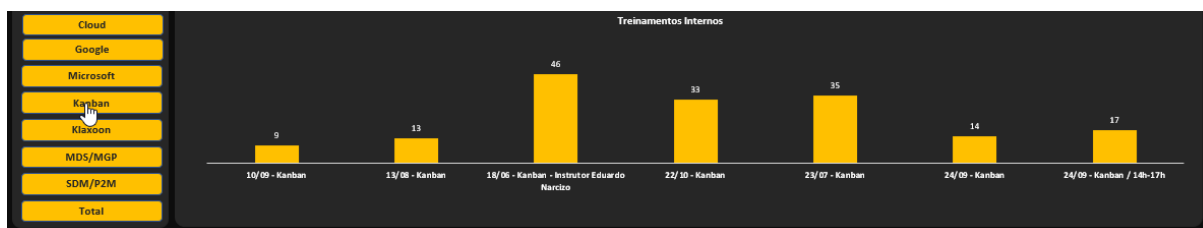
FIGURA 4 - Visibilidade 3.



Fonte: elaborado pelos autores

Os treinamentos internos (sem custo para a empresa), que são ministrados pelos funcionários ou funcionários terceiros, também foram considerados. Estes foram organizados por itens de treinamentos, possibilitando a pesquisa total, ou seja, todos os itens de treinamentos, ou a seleção de um ou um grupo de itens (parte inferior do painel), conforme exemplo abaixo: item de treinamento Kanban (Figura 5). Nessa figura, observa-se que vários treinamentos forma realizados sobre o Kanban, envolvendo vários participantes. A metodologia Kanban se concentra na melhoria contínua da cadeia de valor de um produto ou serviço por meio de um panorama do fluxo de tarefas. A prática é utilizada para traçar o caminho da agilidade empresarial (Anderson; Carmichael, 2016).

FIGURA 5 - Visibilidade 4.



Fonte: elaborado pelos autores

4.5 Avaliação

O processo de validação foi realizado junto à liderança da área de TI da empresa, os quais utilizarão a ferramenta desenvolvida. Foi avaliada a usabilidade do modelo e a relevância das informações para tomadas de decisões, tais como: atendimento ao orçamento anual, informação de atendimento ou não à meta de aprendizagem definida, análise

quantitativa e qualitativa dos treinados ministrados, além de gaps de aprendizagem para serem tratados futuramente.

O modelo foi bem aceito e considerado intuitivo, além disso, todas as informações que até então eram de difícil acesso e de forma segregada, puderam ser acessadas de forma organizada e dinâmica.

4.6 Implantação

A ferramenta foi entregue para uso a partir da validação da liderança de TI. O modelo foi disponibilizado em um local compartilhado, a partir da rede interna da empresa, com acesso restrito às pessoas autorizadas.

5. CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi construir uma solução para análise relacionada às informações referentes aos itens de capacitação do time de TI da empresa, tais como custo total ou por supervisão, categoria (treinamentos interno, externo e *in company*), temas dos treinamentos segmentados pelos pilares de agilidade, cultura, processos e tecnologia, além da quantidade de profissionais treinados. O processo CRISP-DM foi utilizado, ressaltando a importância de cada etapa na construção do projeto de *Business Intelligence*.

O método CRISP-DM inclui vários processos para executar tarefas simples de mineração de dados, pois incentiva as melhores práticas e replicação para outros modelos. E para o tratamento dos dados para os indicadores dos itens de treinamentos analisados nesse estudo, a metodologia se mostrou eficiente. A CRISP-DM fornece uma estrutura uniforme para planejar e gerenciar o modelo e pode ser implementado em qualquer projeto de Data Science, independentemente de seu domínio.

A aplicação do método CRISP-DM foi importante para apontar se as metas previstas para a realização de treinamentos foram alcançadas. Outro fator relevante é o investimento em que tipos de treinamento tem sido realizado, tendo em vista que os treinamentos indicam o que a organização tem almejado para sua atuação no mercado. Nesse sentido, percebe-se que a organização tem buscado a inserção de novas tecnologias, mudanças externas para acelerar o processo e satisfação do cliente.

Por fim, os dados obtidos através da avaliação podem servir de subsídio para tomadas de decisão no que refere aos treinamentos fornecidos pela empresa. A organização pode prever onde investir mais recursos, quais tipos de treinamentos devem ser ministrados e qual melhor forma para otimizar os custos de treinamentos.

As limitações deste estudo se encontra nas restrições de uso do programa somente no departamento de RH. Futuros estudos devem ser realizados para avaliar a aplicação do CRISP-DM em outras áreas da empresa.

REFERÊNCIAS

- ANDERSON, D.J.; CARMICHAEL, A. **Essential kanban condensed**. Londres: Blue hole press. 2016.
- CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. CRISP-DM 1.0: Step-by-step data mining guide. **SPSS inc**, v. 9, n. 13, p. 1-73, 2000.
- HAMAD, R. M. H.; FAYOUMI, M. Scalable agile transformation process (SATP) to convert waterfall project management office into agile project management office. In: **International Arab conference on information technology (ACIT)**. IEEE, p. 1-8, 2018.
- LEAVITT, H. J. Applied organizational change in industry: Structural, technological and humanistic approaches. In: **Handbook of Organizations** (RLE: Organizations). Routledge, p. 1144-1170, 2013.
- KRCMAR, H. **Informations management**. 6. ed. Wiesbaden: Springer Gabler, 2015.
- MARTÍNEZ-PLUMED, F.; CONTRERAS-OCHANDO, L.; FERRI, C.; HERNÁNDEZ-ORALLO, J.; KULL, M.; LACHICHE, N.; FLACH, P. CRISP-DM twenty years later: From data mining processes to data science trajectories. **IEEE Transactions on Knowledge and Data Engineering**, v. 33, n. 8, p. 3048-3061, 2019.
- NEGASH, S. Business intelligence. **Communications of the Association for Information Systems**, v. 13, n. 1, p. 177-195, 2004.
- NORMENT, H. G.; SCHWENKE, T. W.; KOHLBERG, I. **tech/ops**. Nova York: US Army, 1966.
- PASTRANA, B.; TOBÓN, S. Emerging needs of human talent training in leading information technology companies, a socioformative analysis. **World Review of Science, Technology and Sustainable Development**, v. 16, n. 4, p. 303-319, 2020.
- PLOTNIKOVA, V.; DUMAS, M.; MILANI, F. Adaptations of data mining methodologies: a systematic literature review. **Peer Journal of Computer Science**, v. 6, e267, 2020.
- RAMOS, J. L. R.; SEDRAZ SILVA, J. C.; RODRIGUES, R. L.; OLIVEIRA, P. L. S. CRISP-EDM: uma proposta de adaptação do Modelo CRISP-DM para mineração de dados educacionais. In: **Anais... XXXI Simpósio Brasileiro de Informática na Educação**, Natal, RN, 2020.
- SAXENA, G.; AGARWAL, B. B. Data Warehouse Designing: Dimensional Modelling and ER Modelling. **International Journal of Engineering Inventions**, v. 3, p. 28-34, 2014.

SCHNELL, J.; NENTWICH, C.; ENDRES, F.; KOLLEND, A.; DISTEL, F.; KNOCH, T.; REINHART, G. Data mining in lithium-ion battery cell production. **Journal of Power Sources**, v. 413, p. 360-366, 2019.

SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A systematic literature review on applying CRISP-DM process model. **Procedia Computer Science**, v. 181, p. 526-534, 2021.

SHEARER, C. The CRISP-DM Model: The New Blueprint for Data Mining. **Journal of Data Warehousing**, v. 5, n. 4, p. 13-22, 2000.

VERCELLIS, C. **Business intelligence: data mining and optimization for decision making**. Nova Jersey: John Wiley & Sons, 2011.