UNIVERSIDADE FEDERAL DE OURO PRETO - UFOP
ESCOLA DE MINAS - EM
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO, ADMINISTRAÇÃO E ECONOMIA - DEPRO

Pedro Laender

Aplicação de Inteligência de Negócios no Acompanhamento de uma Obra Civil

Ouro Preto
2021

Pedro Laender

Aplicação de Inteligência de Negócios no Acompanhamento de uma Obra Civil

Monografia apresentada ao Curso de Engenharia de Produção da Universidade Federal de Ouro Preto como parte dos requisitos para a obtenção do Grau de Engenheiro de Produção.

Universidade Federal de Ouro Preto

Orientador: Prof. Dr. Helton Cristiano Gomes

Ouro Preto
2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

L158a Laender, Pedro.
Aplicação de Inteligência de Negócios no acompanhamento de uma obra civil. [manuscrito] / Pedro Laender. - 2021.
44 f.: il.: color., gráf..

Orientador: Prof. Dr. Helton Cristiano Gomes.
Monografia (Bacharelado). Universidade Federal de Ouro Preto. Escola de Minas. Graduação em Engenharia de Produção .

1. Construção civil. 2. Inteligência competitiva (Administração). 3. Inteligência de negócios (BI). 4. Mineração de dados (Computação). I. Gomes, Helton Cristiano. II. Universidade Federal de Ouro Preto. III. Título.

CDU 658.5:004.61

Bibliotecário(a) Responsável: Sione Galvão Rodrigues - CRB6 / 2526

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA DE PRODUCAO, ADMINISTRACAO E ECON

FOLHA DE APROVAÇÃO

Pedro Laender

Aplicação de inteligência de negócios no acompanhamento de uma obra civil

Monografia apresentada ao Curso de Engenharia de Produção da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Engenheiro de Produção.

Aprovada em 01 de novembro de 2021.

Membros da banca

Doutor - Helton Cristiano Gomes - Orientador - Universidade Federal de Ouro Preto
Doutor - Aloisio de Castro Gomes Júnior - Universidade Federal de Ouro Preto
Mestrando - Ruan Carlos Silva Menezes Pinheiro - Universidade Federal de Ouro Preto

Helton Cristiano Gomes, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 01/11/2021.

Documento assinado eletronicamente por **Helton Cristiano Gomes**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 01/11/2021, às 09:03, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0239455** e o código CRC **656C609F**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: 3135591540 - www.ufop.br

Dedico este trabalho à minha família por sempre me apoiarem e me manterem em um caminho de integridade e corretude.

Ao Glorioso Ninho do Amor por toda vivência e experiências engrandecedoras na escola de Ouro Preto.

À Grandiosa Escola de Minas e seus professores.

Agradecimentos

Ao Criador, pela vida.

Ao meu orientador Helton Cristiano Gomes, pela atenção, apoio e incentivo neste trabalho.

À Escola de Minas, por toda sua história e contribuição em minha formação individual e profissional.

Aos professores do curso de Engenharia de Produção pelos ensinamentos.

À vida republicana de Ouro Preto.

Aos irmãos republicanos.

"O caráter de um homem é seu destino.
O caráter de um homem é sua divindade guardiã."
Heráclito

Resumo

A Inteligência de Negócios está sendo aplicada por um grande número de organizações dos mais variados setores, sendo considerada prioridade para as empresas se manterem competitivas no mercado. O setor da construção civil possui grande expressão na economia nacional, mas apesar disso, apresenta uma lacuna considerável na busca por novas tecnologias. O presente trabalho buscou evidenciar possíveis benefícios ao se utilizar ferramentas de BI no acompanhamento de uma obra civil. Realizando uma análise descritiva dos dados obtidos de uma obra civil, criou-se um relatório dinâmico e interativo buscando exibir as principais informações relacionadas ao processo de gerenciamento do projeto.

Palavras-chave: Inteligência de Negócios; Construção civil; Power BI; Mineração de dados.

Abstract

Business Intelligence is being applied by a large number of organizations from many sectors, being considered a priority for companies to remain competitive in the market. The construction sector has great expression in the national economy, but despite this fact, it has a huge gap in the search for new technologies. The present work sought to highlight possible benefits by using BI tools in the follow-up of a construction. Performing a descriptive analysis of the data obtained from a construction, a dynamic and interactive dashboard was created, seeking to display the main informations related to the process of managing the project.

Keywords: Business Intelligence; Construction; Power BI; Data Mining.

Lista de abreviaturas e siglas

BI	Inteligência de Negócios
DW	Data Warehouse
OLAP	Processamento Analítico Online
DDS	Sistemas de Apoio à Decisão
MIS	Sistemas de Informação Gerencial
BPM	Gestão de Desempenho de Negócios
KDD	Descoberta de Conhecimento nas Bases de Dados

Lista de ilustrações

Sumário

1 Introdução

A construção civil possui grande expressão na economia nacional. De acordo com o Instituto Brasileiro de Geografia e Estatística (2021), a receita proveniente desse setor corresponde a aproximadamente 25% de toda receita produzida pela indústria brasileira, além de ter apresentado um contingente de pessoas ocupadas de aproximadamente 6,0 milhões de brasileiros no primeiro trimestre de 2021. Esse número representa 6,0% da população total ativa desse período.

A indústria como um todo, incluindo as empresas de construção civil, possui uma lacuna considerável na busca por tecnologias inovadoras e aperfeiçoamento de seus protocolos de gerenciamento. É oportuno, portanto, uma análise dos potenciais benefícios obtidos com a implantação de tecnologias que possam aprimorar o desempenho dos processos de gestão dessas empresas, conferindo celeridade e um maior índice de acertos nas tomadas de decisão críticas para o bom funcionamento de seus processos.

Nesse aspecto, a Inteligência de Negócios (BI), amplamente difundida entre as grandes corporações que vêm se mantendo competitivas frente a um mercado globalizado e em rápida e constante mudança (ISIK; JONES; SIDOROVA, 2013), emerge como uma solução confiável para criar conhecimento e inteligencia úteis a serem aplicados também no setor da construção civil.

A competição acirrada nos mundo dos negócios, tem feito as pessoas perceberem a importância da alocação adequada dos seus recursos e de ter a informação correta no momento correto para o processo de tomada de decisão e para o bom funcionamento da organização. Todos os dias as empresas tem armazenado uma enorme quantidade dos mais variados tipos de dados, tanto internos quanto externos. Cria-se então a necessidade de se processar e utilizar essas informações de forma precisa para o sucesso da organização.

Desde o início da utilização de computadores para auxiliar no gerenciamento das organizações, podendo citar os MIS's (Sistemas de Informação Gerencial) e os DDS's (Sistemas de Apoio à decisão), as empresas vem percebendo os enormes benefícios que essas tecnologias são capazes de prover, ao conseguirem lidar com uma enorme quantidade de dados complexos. Nos último anos, o BI tem se desenvolvido como uma poderosa ferramenta para ajudar analistas a encontrar tendências de mercado e monitorar riscos, além de estar mudando o ambiente de negócios para um modelo com tomadas de decisões mais rápidas e racionais. Esses fatores potencializam a competitividade e trazem grandes benefícios às organizações.

Na construção civil, esse problema se confirma. O excesso de informação coletada, quando não processada da maneira correta, acaba por gerar desinformação. Ao se aplicar ferramentas de BI nesse setor, as empresas conseguem ter uma visão mais completa dos seus projetos de implantação, além de maior flexibilidade, agilidade e precisão para

construir suas análises e poderem relacionar dados que até então pareciam não apresentar conexão direta.

1.1 Objetivos

O objetivo deste trabalho é verificar possíveis benefícios gerados a partir da aplicação de programas e técnicas da Inteligência de Negócios no acompanhamento de uma obra civil.

1.1.1 Específicos

- Estudar os dados coletados de uma obra civil.
- Utilizar o programa Power BI para criar interfaces interativas, que permitam a visualização de dados importantes sobre a obra de maneira descomplicada.
- Gerar informações úteis sobre a obra que auxiliem nas tomadas de decisão.

1.2 Estrutura do Trabalho

O presente trabalho é apresentado com cinco diferentes tópicos, para um melhor entendimento. O primeiro aborda a formulação do problema a ser respondido, a justificativa da realização do estudo e os objetivos gerais e específicos.

O segundo capítulo refere-se à revisão bibliográfica, com a apresentação de conceitos fundamentais e base teórica para o entendimento do assunto em questão. Serão abordados conceitos relacionados à Inteligencia de Negócios, englobando algumas definições do termo, uma breve entendimento de sua origem, os tipos de análises que podem ser realizadas, além de apresentar a arquitetura do BI, descrevendo as principais tarefas e técnicas utilizadas.

O terceiro capítulo refere-se à metodologia aplicada para a obtenção e análise dos resultados obtidos do objeto de estudo teste trabalho.

O quarto capítulo mostra os resultados e discussões envolvendo as informações mais relevantes extraídas dos dados.

O quinto e último capítulo apresenta as conclusões fazendo referência aos resultados e estudos realizados.

2 Referencial teórico

"Sistemas de BI combinam dados históricos e operacionais com ferramentas analíticas para apresentar informações úteis e competitivas aos gestores e tomadores de decisão nos negócios."(KHAN; QUADRI, 2012). Com as rápidas mudanças e a globalização dos mercados as empresas estão sempre buscando novas ferramentas e tecnologias que as ajudem a se manter competitivas frente aos novos desafios. O BI tem como objetivo transformar dados em conhecimento com o propósito de fornecer às pessoas as informações que elas necessitam para realizar seu trabalho (ARGOTTE; MEJIA-LAVALLE; SOSA, 2009). Por isso o BI se tornou um fator crítico para a competitividade em muitas organizações e tem sido classificado como uma das prioridades dos grandes executivos nos últimos anos (ISIK; JONES; SIDOROVA, 2013). Neste capítulo serão apresentadas definições de BI, assim como sua arquitetura e algumas ferramentas utilizadas no processo de transformar dados crus em informações relevantes.

2.1 Definições de BI

Segundo Sharda, Delen e Turban (2019), Inteligência de Negócios (BI) é um termo amplo que combina arquiteturas, bases de dados, ferramentas analíticas, aplicativos e metodologias com o objetivo de possibilitar acesso interativo a dados e permitir sua manipulação para fornecer aos gestores, análises que os auxiliem a tomar decisões mais apropriadas. O processo de BI baseia-se na transformação de dados em informações, depois em decisões e por fim em ações. Dentre as funcionalidade do BI, podemos destacar extração dinâmica de relatórios multidimensionais, geração de previsões, análise de tendências, aprofundamento em detalhes, acesso a status e fatores cruciais de sucesso e funcionalidades de inteligência artificial.

Para Lönnqvist e Pirttimäki (2006), o termo BI pode ser utilizado para se referir a informações relevantes e conhecimentos que descrevam o ambiente do negócio, sua organização e situação em relação ao mercado, consumidores, competidores e questões econômicas, ou a um processo sistemático e organizado pelo qual organizações adquirem, analisam e distribuem informações consideradas relevantes para o negócio e tomadas de decisão.

Os mesmos autores ainda complementam essa ideia ao afirmarem que o BI aborda os mesmo velhos problemas gerenciais de sempre, que consistem em analisar o complexo ambiente dos negócios a fim de tomar melhores decisões. Sendo assim, a finalidade do BI é auxiliar na identificação e processamento da enorme quantidade de informações a respeito do ambiente de negócios em conhecimento e inteligencia úteis para a gerência.

Já para Khan e Quadri (2012), BI é um conceito utilizado para descrever a análise de dados coletados com a intenção de auxiliar equipes de tomadores de decisão a compreen-

der melhor as operações das empresas, possibilitando melhores decisões nos negócios. É utilizado para melhorar a qualidade e oportunismo das informações permitindo um melhor entendimento por partes dos gestores a respeito da posição de sua organização em comparação com os competidores. As tecnologias do BI ajudam as companhias a analisar tanto as mudanças em relação ao mercado quanto aos padrões de comportamento de seus consumidores, auxiliando os gestores e analistas a determinar novas estratégias para se adequar à essas mudanças. BI pode ser apresentado como uma arquitetura, ferramenta, tecnologia ou sistema que coleta e armazena dados que são investigados e analisados para gerar informações e conhecimento que, em última instancia permitem uma melhor tomada de decisão por parte das organizações.

Apesar das muitas definições, em geral as ferramentas de BI correspondem a uma ou várias das categorias listadas a seguir: ferramentas de *Data Warehouse* (DW), que são repositórios de dados com capacidade de armazenar grandes volumes de dados e destinados a suportar os relatórios empresariais e o processo decisório, os quais permitem a extração, transformação e carregamento (ETL) de grandes volumes de dados para serem processados e transformados em informação; ferramentas de mineração de dados, cujo objetivo é colher e processar informações presentes nas bases de dados para gerar novas informações uteis nos processos gerenciais; sistemas de Processamento Analítico Online (OLAP), os quais fornecem um ambiente dinâmico, multidimensional e consolidado para análise de dados; ferramentas de tomada de decisão que auxiliam as organizações em processos decisórios; e sistemas de gestão de conhecimento focados na preservação e disseminação de informações e conhecimentos (HARISON, 2012).

2.2 Origem

A Inteligência de Negócios tem sua origem na década de 70 quando Scott-Morton definiu pela primeira vez os principais conceitos dos Sistemas de Apoio à Decisão (DDS) diferenciando-os assim dos Sistemas de Informação Gerencial (MIS) (SHARDA; DELEN; TURBAN, 2019). Para Keen e Scott-Morton (1978 apud (SHARDA; DELEN; TURBAN, 2019), p.12) DDSs são sistemas de apoio computadorizados que complementam os recursos intelectuais dos tomadores de decisões gerenciais com as capacidades do computador para melhorar a qualidade das decisões.

Com o passar dos anos, a crescente necessidade de se obter vantagem competitiva no mercado fez surgirem novos sistemas e recursos para auxiliar os indivíduos a tomarem decisões mais acertadas e com maior rapidez. Na virada do milênio os DDSs baseados em *Data Warehouses* passaram a ser chamados de *Business Intelligence* (BI). (SHARDA; DELEN; TURBAN, 2019).

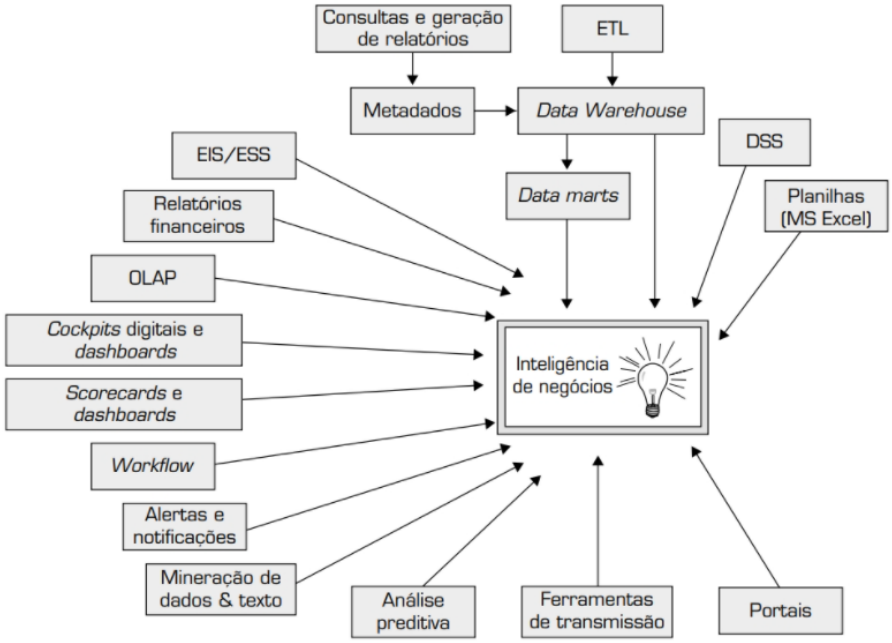

Figura 1 – Evolução da Inteligência de Negócios (SHARDA; DELEN; TURBAN, 2019).

2.3 Tipos de análise de dados

Segundo Sharda, Delen e Turban (2019) a análise de dados pode ser dividida em três níveis, que representam 3 etapas de certa forma independentes, porém havendo uma certa sobreposição entre elas pois a aplicação de um tipo leva a outro. Estes níveis são:

Análise de dados Descritiva: segundo Sharda, Delen e Turban (2019) e Reis e Reis (2002) é a fase inicial do processo de análise de dados, utilizada para conhecer o que está ocorrendo na organização no que tange a entender tendências e causas subjacentes das ocorrências. Para tal é necessário a consolidação de fontes de dados, ou seja, organizar, resumir e descrever os aspectos importantes de um conjunto de características de forma a permitir a extração e análise apropriadas de relatórios. Nessa etapa são utilizadas ferramentas mais voltadas para a visualização como gráficos, tabelas e medidas de síntese como porcentagens, índices e médias. Outra característica desse nível é permitir a identificação de possíveis anomalias nos dados.

Análise de dados Preditiva: é o nível em que a organização busca prever o que acontecerá e porquê aquilo é mais provável de ocorrer no futuro, ou seja, o intuito desta etapa é realizar projeções precisas de eventos futuros e resultados finais. Para tal são utilizados inúmeras técnicas de estatística e mineração de dados.

Análise de dados Prescritiva: é a terceira etapa da análise de dados e de acordo com Sharda, Delen e Turban (2019) tem como objetivo direcionar as decisões da organização para garantir o melhor desempenho possível, chegando em uma decisão ou

recomendação para uma ação específica. Para se obter sucesso nesse nível utiliza-se programas de otimização e simulação, além de modelos de decisão e sistemas especialistas.

2.4 Mensurando BI

A medição da performance dos negócios é uma atividade de suma importância para qualquer organização e vem sendo aprimorada ao longo dos anos (LöNNQVIST; PIRTTIMäKI, 2006). Essa medição pode ser utilizada para vários propósitos: tomada de decisão, controle, orientação, educação e aprendizado e comunicação externa (SIMONS 2000 apud (LöNNQVIST; PIRTTIMäKI, 2006)). Para o BI essa importância se confirma, mas é algo difícil de se atingir (HANNULA; PIRTTIMäKI, 2003).

Para Lönnqvist e Pirttimäki (2006) e Pirttimäki, Lönnqvist e Karjaluoto (2006) mensurar o BI tem dois propósitos. O primeiro e mais comum na literatura reside na necessidade de se provar que os investimentos nessa área tem o seu retorno. O gerente da área precisa medir os retornos que o BI traz para a empresa a fim de justificar a existência do seu departamento, uma vez que medidas concretas e confiáveis aumentam a credibilidade do BI entre as empresas. O segundo propósito é ajudar a gerenciar o processo de BI dentro da empresa para garantir que seja eficiente e que o seu produto atenda às necessidades do usuário, pois a implementação do BI pode ter um custo alto se as informações geradas não corresponderem às necessidades da empresa. Ademais, os responsáveis por essa medição normalmente são os próprios profissionais de BI da empresa, permitindo assim, que eles utilizem estas informações para continuamente melhorar os produtos e serviços do BI.

2.4.1 Mensurando o valor do BI

Valor é um conceito difícil de se trabalhar. Na literatura é normalmente analisado pelo ponto de vista da empresa, que requer que seus investimento gerem um mínimo de retorno, ou do usuário, por meio da sua percepção da utilidade do BI. Em resumo para poder avaliar o valor do BI deve-se analisar os custos de sua aplicação e os benefícios gerados. Os custos são fáceis de se calcular, para tal deve-se levar em conta um investimento inicial e os custos de operação como custos do trabalho, de compra de informações, materiais dentre outros (DAVISON, 2001; LöNNQVIST; PIRTTIMäKI, 2006).

O cálculo dos benefícios gerados por sua vez, é uma tarefa mais complicada, uma vez que estes benefícios são muitas vezes intangíveis como aprimoramento da tomada de decisão, oportunismo e melhor qualidade de informação. Herring (1996 apud (LöNNQVIST; PIRTTIMäKI, 2006)) identificou outra forma de medir a efetividade do BI, que seria analisando economia de tempo e de custos, prevenção de custos e aumento de receita. Estas porém, também são medidas subjetivas pois é difícil identificar com precisão quais destes aspectos tem seus resultados vindos da aplicação do BI. Davison (2001) apresenta ainda

outra forma subjetiva de medir a efetividade do BI ao se analisar a satisfação percebida do usuário, i.e. verificar com o usuário como o produto do BI tem auxiliado nos quesitos de entrega de informação em momento oportuno e confiança para tomada de decisão por exemplo. Porém, métodos subjetivos não geram evidência do valor monetário criado pela utilização de BI (LöNNQVIST; PIRTTIMäKI, 2006; PIRTTIMäKI; LöNNQVIST; KARJALUOTO, 2006).

2.4.2 Mensurando para administrar o processo de BI

Ao se tratar da mensuração para administrar o processo do BI, as medidas de efetividade e as subjetivas abordadas na seção anterior ainda se mostram válidas, porém, o foco dessa área de medição deve ser voltado para o profissional de BI e devem ser aplicadas para a melhoria continua do processo. Nesse aspecto os esforços se concentram em produzir inteligência com valor e de forma efetiva para suprir as necessidades específicas de cada usuário. Alguns fatores importantes a serem destacados nesse quesito são a eficiência dos usuário de BI, alocação efetiva dos recursos disponíveis, qualidade do produto obtido e a satisfação do usuário quanto aos produtos e serviços gerados pelo BI. A medição dessa satisfação pode ser realizada ao analisar a qualidade, relevância, oportunismo, precisão e capacidade de ação gerada pelas informações produzidas pelo BI. Outra forma de realizar essa mensuração seria ao medir a proporção dos gerentes que utilizam as ferramentas de BI e quão frequentemente elas são acessadas. (LöNNQVIST; PIRTTIMäKI, 2006).

2.5 Arquitetura

Para Khan e Quadri (2012) a arquitetura do BI pode ser divida em 3 partes: coleta, armazenamento e acesso e análise dos dados. Os dados são coletados de fontes internas, base de dados operacional da organização e banco de dados, ou externa, dados dos consumidores, fornecedores, agências do governo e competidores. Os dados são então armazenados nos DW depois de passarem pelo processo de extração, transformação e carregamento (ETL), para em seguida serem analisados para auxiliarem nas tomadas de decisão.

Sharda, Delen e Turban (2019) complementam essa ideia ao afirmarem que a arquitetura do BI é composta de 4 fatores: um DW com os dados coletados; as ferramentas utilizadas para minerar, manipular e analisar dados que foram denominadas análise de negócios; a Gestão de Desempenho de Negócios (BPM) para monitorar e analisar desempenhos; e uma interface do usuário como um painel interativo por exemplo.

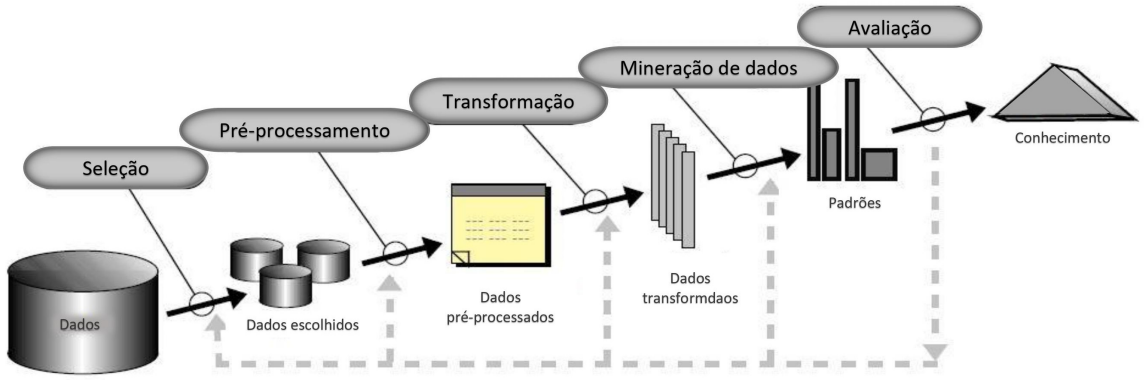

Figura 2 – Visão geral do processo de coleta, armazenamento e análise dos dados (FAYYAD; PIATETSKY-SHAPIR; SMYTH, 1996).

2.6 Os dados

Ter conhecimento sobre os tipos de dados com que se está trabalhando é fundamental para a etapa de pré-processamento e para poder escolher os métodos mais adequados de análise e visualização. Os dados podem ser classificados como quantitativos ou qualitativos. Os dados qualitativos são representados por valores nominais (categóricos), ou seja, cada valor representa uma categoria, código ou estado. Os quantitativos são representados por valores numéricos e podem ainda ser discretos ou contínuos (HAN; KAMBER; PEI, 2011).

Antes de poder trabalhar com os dados é necessário verificar sua integridade. Muitos dos dados brutos contidos em bancos de dados não são processados, são incompletos e possuem inconsistências. Alguns aspectos indesejáveis que um banco de dados pode conter são: campos obsoletos ou redundantes, valores ausentes, "outliers" (valores atípicos), dados em um formato não adequado para modelos de mineração de dados e valores não consistentes com a política ou o bom senso. Para serem úteis para fins de mineração de dados, os bancos de dados precisam passar por uma etapa de preparação, também conhecida como pré-processamento, cujo objetivo principal é minimizar o "lixo" que fica no modelo para minimizar a quantidade de "lixo" que o modelo fornece. Ademais, as técnicas de processamento de dados, quando aplicadas antes da mineração, podem melhorar substancialmente a qualidade geral dos padrões extraídos e o tempo necessário para a mineração (HAN; KAMBER; PEI, 2011). No pré-processamento, os dados passam pelas etapas de limpeza, integração, redução e transformação (LAROSE, 2005)

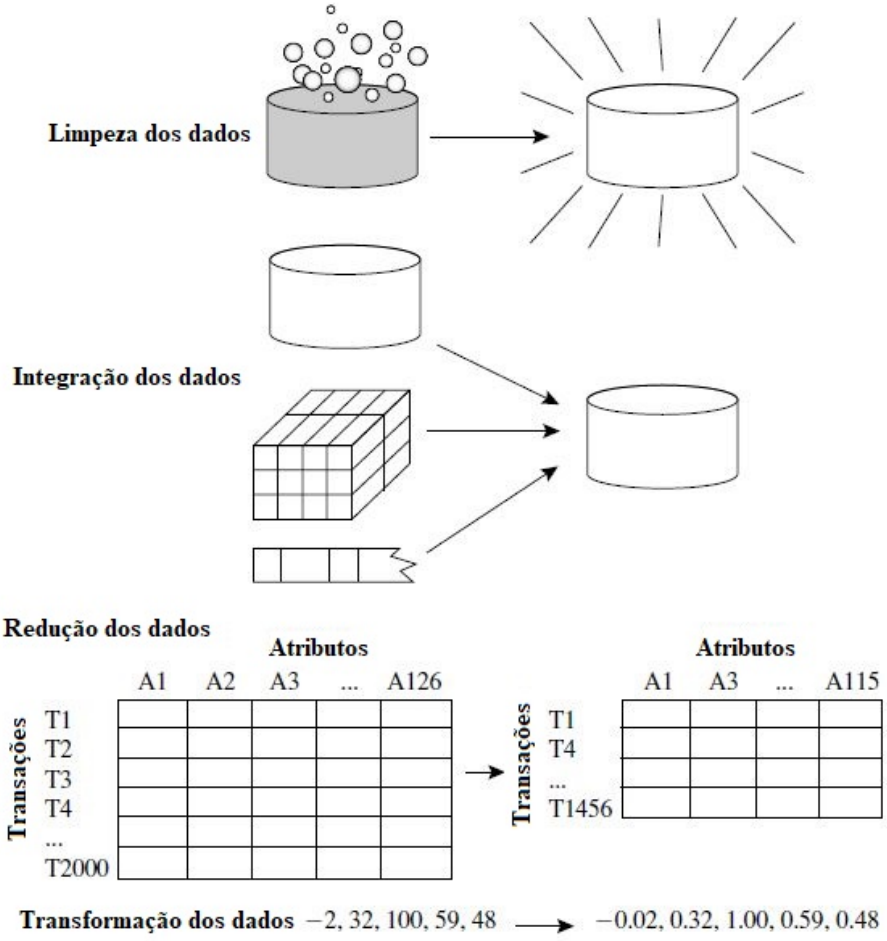

Figura 3 – Etapas de pré-processamento dos dados (HAN; KAMBER; PEI, 2011).

Limpeza dos dados: frequentemente, os dados são encontrados com diversas inconsistências: registros incompletos, valores errados e dados inconsistentes. A etapa de limpeza de dados visa suprimir estes problemas para que eles não tenham influência nos resultados. Se usuários acreditam que os dados estão sujos, é improvável que eles confiem nos resultados de qualquer mineração de dados que tenha sido aplicada. Além disso, dados sujos podem causar confusão para o procedimento de mineração, resultando em saída não confiável (HAN; KAMBER; PEI, 2011). A limpeza de dados ajuda as organizações na questão de criar uma visão lógica unificada da grande variedade de dados e bancos de dados que possuem de forma a resolver os problemas de mapeamento dados criando uma única convenção de nomenclatura, representada uniformemente e tratando dados inválidos ou perdidos, além de inconsistências e erros quando possível (FAYYAD; PIATETSKY-SHAPIR; SMYTH, 1996).

Integração dos dados: a heterogeneidade das fontes dos dados muitas vezes requer que eles passem pelo processo de integração, que combina dados de fontes múltiplas para se obter um repositório coerente, único e consistente. A integração cuidadosa

pode ajudar a reduzir e evitar redundâncias, inconsistências e valores conflitantes (categorias diferentes para os mesmos valores, chaves divergentes, regras diferentes para os mesmos dados, entre outros). No conjunto de dados resultante, o que pode ajudar a melhorar a precisão e a velocidade do processo de mineração de dados (HAN; KAMBER; PEI, 2011)

Redução dos dados: a redução de dados obtém uma representação reduzida do conjunto de dados que é muito menor em volume, mas produz os mesmos, ou quase os mesmos resultados analíticos. As estratégias de redução dos dados incluem redução de dimensionalidade e de numerosidade. A redução da dimensionalidade é o processo de redução do número de variáveis ou atributos aleatórios em consideração. As técnicas de redução de numerosidade substituem o volume de dados original por formas alternativas e menores de representação de dados. Na compressão de dados, as transformações são aplicadas de forma a obter uma representação reduzida ou comprimida dos dados originais (HAN; KAMBER; PEI, 2011).

Transformação dos dados: nessa etapa os dados são transformados para facilitar seu processamento. Não existe um critério único para transformação dos dados e diversas técnicas podem ser usadas de acordo com os objetivos pretendidos. Algumas técnicas empregadas nessa etapa são: normalização (colocar as variáveis em uma mesma escala), suavização (eliminar valores inconsistentes), agregação (agregar os valores em faixas resumidas), criação de novos atributos e transformação de variáveis numéricas em valores categóricos (SHARDA; DELEN; TURBAN, 2019).

Todas as etapas do processo de descoberta de informações de bases de dados como a preparação, seleção, limpeza dos dados e interpretação apropriada dos resultados da mineração são essenciais para garantir a utilidade dos conhecimentos gerados a partir dos dados. Aplicar as ferramentas de mineração de dados de maneira descuidada pode ser perigoso, levando a descobertas sem sentido e padrões inválidos (FAYYAD; PIATETSKY-SHAPIR; SMYTH, 1996).

Uma vez tratados os dados devem então ser armazenados em *Data Warehouses*. Um DW é um repositório de dados que generalizam e consolidam dados geralmente modelados por uma estrutura de dados multidimensional, chamada cubo de dados, em que cada dimensão corresponde a um atributo ou conjunto de atributos no esquema, e cada célula armazena o valor de alguma medida agregada, como contagem ou soma. Um cubo de dados fornece uma visão multidimensional dos dados e permite a pré-computação e o acesso rápido aos dados resumidos. Eles apoiam o processamento de informações, fornecendo uma plataforma sólida de dados históricos consolidados para análise. Em suma, é um local de armazenamento semanticamente consistente que serve como um implementação física de um modelo de dados de suporte à decisão. Ele armazena as informações que uma empresa precisa para tomar decisões estratégicas e é frequentemente visualizado como

uma arquitetura, construída integrando dados de várias fontes heterogêneas para apoiar consultas estruturadas ou sob demanda, relatórios analíticos e tomada de decisão. Além disso, os DW fornecem ferramentas de Processamento Analítico Online (OLAP) para a análise interativa de dados multidimensionais, o que aumenta a eficácia da mineração de dados. Conseqüentemente, o DW tornou-se uma plataforma cada vez mais importante para análise de dados e processamento analítico online OLAP fornecendo uma plataforma eficaz para mineração de dados. Portanto, armazenamento de dados e OLAP constituem uma etapa essencial no processo de descoberta de conhecimento (HAN; KAMBER; PEI, 2011).

O OLAP provê ao usuário os meios de explorar e analisar grandes quantidade de dados, envolvendo cálculos complexos, seus relacionamentos e apresenta os resultados em diferentes perspectivas. A visão multidimensional fornecida pelo OLAP provê acesso aos dados de forma rápida e flexível. Os recursos utilizados para fornecer essa visão incluem: possibilidade de agrupar os dados de forma generalizada ou revelar uma maior gama de detalhes e navegar nas dimensões mudando a perspectiva de visualização dos dados, além de análises complexas como séries temporais, previsões e modelagem estatística (KHAN; QUADRI, 2012).

Muitas outras funções de mineração de dados, como associação, classificação, predição e agrupamento podem ser integrados com operações OLAP para aprimorar a mineração interativa de conhecimento em vários níveis de abstração (HAN; KAMBER; PEI, 2011).

2.7 Mineração de Dados

Para Fayyad, Piatetsky-Shapir e Smyth (1996) a mineração de dados é uma etapa no processo de KDD (*Knowledge Discovery in Databases* ou Descoberta de Conhecimento nas Bases de Dados) que consiste em aplicar análise de dados e descobrir algoritmos que produzem uma enumeração particular de padrões dos dados alvo. De acordo com Larose (2005), a mineração de dados é o processo de descoberta de novas correlações, padrões e tendências significativas ao analisar grandes quantidades de dados armazenados em repositórios, utilizando o reconhecimento de padrões tecnologias, bem como técnicas estatísticas e matemáticas, que está se tornando mais difundido a cada dia, porque capacita empresas a descobrir padrões lucrativos e tendências em seus bancos de dados existentes. Já para Sharda, Delen e Turban (2019) a mineração de dados é uma tecnologia facilitadora para a análise de negócios com o intuito de desenvolver informações ou conhecimentos práticos, a partir dos dados armazenados pelas organizações, ajudando a criar uma visão holística do ambiente em que a empresa está inserida. Em termos técnicos, mineração de dados é um processo que emprega técnicas estatísticas, matemática e de inteligência artificial para extrair e identificar informações uteis e conhecimentos a partir de vastos conjuntos de dados. Khan e Quadri (2012) descrevem a mineração de dados como sendo a utilização de uma variedade de técnicas para identificar informações ou conhecimentos úteis para

tomada de decisão em uma grande quantidade de dados. A ideia da mineração de dados é extrair valiosas informações de lugares inesperados, ao utilizar programas e técnicas para identificar relações e padrões globais escondidos em uma vasta quantidade nos dados.

O processo de mineração de dados envolve repetidas aplicações iterativas de métodos de mineração de dados e pode ter seu foco em dois objetivos distintos: verificação e descoberta. Quando se tratando da verificação, o sistema está limitado em verificar as hipóteses do usuário, enquanto na descoberta o sistema procura por novos padrões. A verificação ainda pode ser subdivida em dois objetivos, previsão e descrição. A previsão envolve a utilização de algumas variáveis ou campos da base de dados para prever valores futuros ou desconhecidos de outras variáveis de interesse, ou seja, na previsão procura-se prever o comportamento de entidades enquanto a descrição visa descrever padrões presentes nos dados de uma forma compreensível para o usuário (FAYYAD; PIATETSKY-SHAPIR; SMYTH, 1996).

Um erro que as companhias cometem é pensar que a mineração de dados representa um conjunto de ferramentas isoladas a serem aplicadas de maneira superficial sobre um conjunto de dados. Ao tentarem implementar a mineração de dados dessa forma, as organizações diminuem muito suas chances de sucesso, pois a mineração de dados deve ser vista com uma parte de um processo. A implantação de modelos de mineração de dados geralmente representa uma despesa de capital e investimento em a parte da empresa. Se os modelos em questão forem inválidos, o tempo da empresa e dinheiro é desperdiçado (LAROSE, 2005).

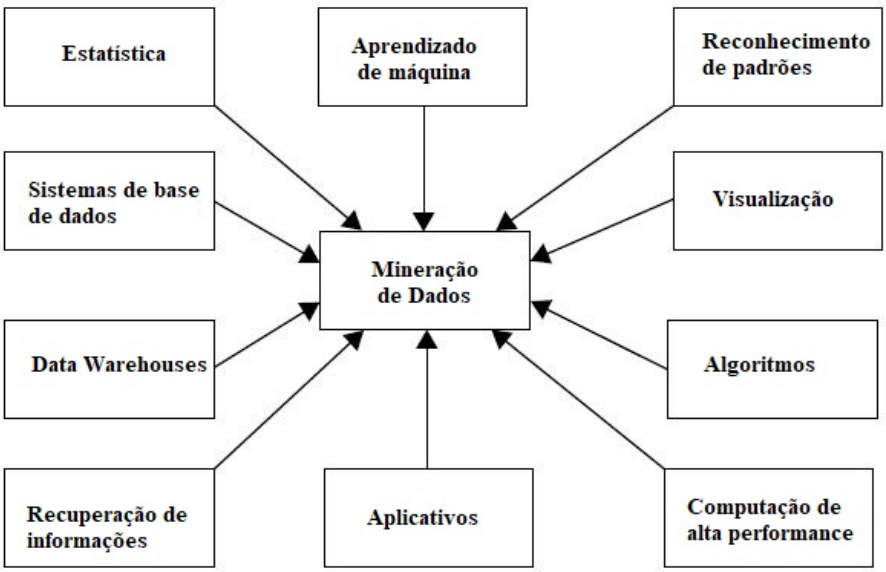

Figura 4 – Mineração de dados adota técnicas de vários domínios (HAN; KAMBER; PEI, 2011).

2.7.1 Tarefas

Existem várias funcionalidades de mineração de dados que são utilizadas para especificar os tipos de padrões a serem encontrados nas tarefas de mineração. Em geral, tais tarefas podem ser classificadas em duas categorias: descritivas e preditivas. Na mineração descritiva, as tarefas caracterizam as propriedades dos dados em um conjunto de dados de destino. Tarefas de mineração preditivas visam realizar indução nos dados atuais para fazer previsões (HAN; KAMBER; PEI, 2011). As tarefas mais comuns são:

Descrição (Description): é a tarefa utilizada para descrever os padrões e tendências extraídos dos dados. A descrição normalmente fornece possíveis explicações para os resultados encontrados. Os resultados dos modelos devem ser o mais transparente possíveis, descrevendo os resultados de maneira clara e intuitiva. Alguns métodos são mais adequados que outros em se tratando de interpretações transparentes portanto deve ser escolhido de maneira cautelosa. A tarefa de descrição é muito utilizada em conjunto com técnicas de análise exploratória de dados, demonstrando os padrões e tendências graficamente, o que facilita a compreensão pelo ser humano (LAROSE, 2005).

Classificação (Classification): É uma das tarefas mais comuns e seu o objetivo é identificar a qual categoria um determinado registro pertence tendo em vista uma variável alvo. Nesta tarefa, o modelo analisa o conjunto de registros fornecido, sendo que cada registro já contém uma indicação de sua categoria a fim de "aprender" a classificar novos registros (aprendizagem supervisionada). Por exemplo, suponhamos que se deseja classificar pessoas em categorias de renda alta, média ou baixa, ou seja, a variável alvo neste caso é a renda. O algoritmo então, analisa vários registros contendo o sexo, idade, profissão e renda de pessoas para então poder classificar novos registros que não contenham a informação de renda, em qual categoria essa pessoa se encaixa. Alguns métodos comumente utilizados para classificação são arvores de decisão, redes neurais e k-ésimo vizinho mais próximo (LAROSE, 2005). As tarefas de classificação podem ser utilizas para:

Determinar se uma transação de cartão de crédito é fraudulenta.

Avaliar se uma aplicação de hipoteca tem risco alto ou baixo.

Determinar se um testamento foi realmente escrito pelo falecido ou se foi fraudado.

Regressão (Regression): É similar à classificação, porém é utilizada para modelar registros de valor numérico, não categórico. Esses modelos visam estimar o valor de novas observações analisando os valores de variáveis já conhecidos. A classificação e a regressão podem precisar ser precedidas por uma análise de relevância, que tenta identificar atributos que são significativamente relevantes para a classificação

e processo de regressão. Esses atributos serão selecionados para a os processos de classificação e regressão enquanto outros, que são irrelevantes, podem ser excluídos da consideração (HAN; KAMBER; PEI, 2011). A regressão pode ser utilizada para analisar situações como:

Estimar a quantidade de dinheiro gasto por uma família de 4 pessoas no período de volta às aulas.

Estimar a probabilidade de um paciente sobreviver dado os resultados de testes diagnósticos.

Prever a demanda de consumo de um produto em função das despesas em publicidade.

Predição (Prediction): A predição atraiu atenção considerável, dadas as implicações potenciais de uma previsão bem-sucedida em um contexto de negócios. Existem dois tipos principais de previsões: pode-se tentar prever alguns valores de dados indisponíveis ou tendências pendentes, ou prever um rótulo de classe para alguns dados. Este último está vinculado à classificação. Uma vez que um modelo de classificação é construído com base em um conjunto de treinamento, o rótulo de classe de um objeto pode ser previsto com base nos valores de atributo do objeto e nos valores de atributo das classes. A predição é, no entanto, mais frequentemente referida à previsão de valores numéricos ausentes ou tendências de aumento e diminuição nos dados relacionados ao tempo. A ideia principal é usar um grande número de valores passados para considerar os valores futuros prováveis (ALASPURKAR, 2013). Alguns exemplos dessa tarefa incluem:

Prever o preço de ações três meses no futuro.

Prever o incremento na porcentagem de mortes no trânsito se o limite de velocidade for aumentado.

Prever o vencedor de um campeonato baseando-se em análises estatísticas dos times.

Agrupamento (Clustering): Se refere a agrupar os registros e observações em classes de objetos similar. Um agrupamento é uma coleção de registros similares entre si e diferentes dos registros em outros agrupamentos. Essa tarefa difere da classificação uma vez que não possui uma variável alvo. Ao invés de classificar, prever ou estimar o valor de uma variável alvo, o agrupamento visa segmentar os dados em subgrupos relativamente homogêneos nas quais a similaridade dos dados dentro do agrupamento é maximizada a similaridade fora é minimizada. O agrupamento é normalmente utilizado como uma etapa preliminar no processo de mineração de dados, com os resultados dessa tarefa sendo utilizados como dados de entrada para outras técnicas e métodos (LAROSE, 2005). Na Inteligência de Negócios, o agrupamento pode ser utilizado para organizar um grande número de clientes em grupos, onde os clientes

dentro de um grupo compartilham fortes características semelhantes. Isso facilita o desenvolvimento de estratégias de negócios para clientes aprimorando a gestão de relacionamento (HAN; KAMBER; PEI, 2011). Alguns exemplos de tarefas de agrupamento são:

Segmentação de mercado para um produto de nicho

Em auditorias, para separar comportamentos suspeitos

Como ferramenta de redução quando os dados apresentam centenas de atributos

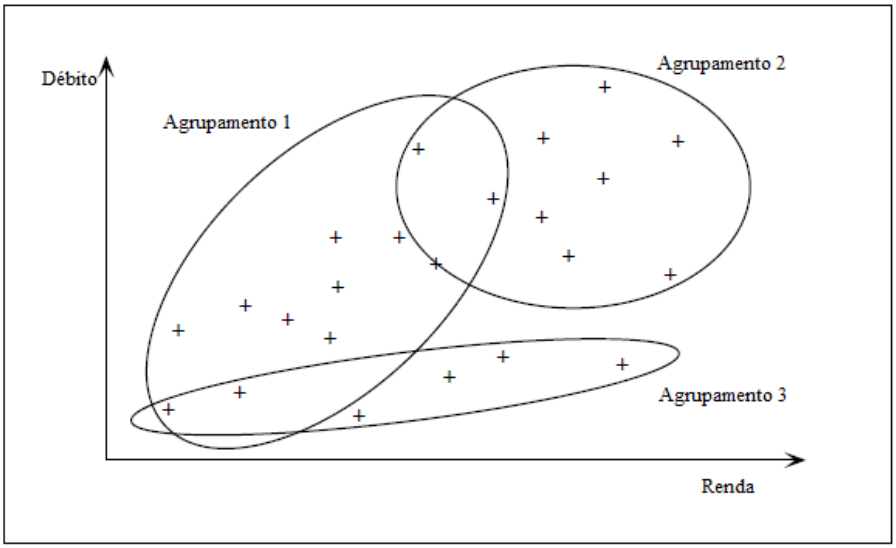

Figura 5 – Exemplo de agrupamento (FAYYAD; PIATETSKY-SHAPIR; SMYTH, 1996).

Associação (Association): é uma tarefa que busca encontrar quais atributos se relacionam. Muito utilizada no mundo dos negócios, também conhecida como análise de afinidade ou análise de cesta de mercado (*Market Basket*), a associação busca descobrir as regras para quantificar a relação entre dois ou mais atributos seguindo a regra "Se atributo X, então atributo Y". Um exemplo para clarificar esse conceito é encontrar qual a porcentagem de pessoas que vão em um supermercado e compram ao mesmo tempo fralda e cerveja (LAROSE, 2005). Exemplos de associação são:

Investigar a porcentagem de usuário de determinada companhia telefônica que responderiam positivamente a uma oferta de melhoria de plano.

Examinar a proporção de crianças cujos pais leem ara ela que se tonam bons leitores.

Determinar a proporção de casos em que um novo medicamento irá causar efeitos colaterais.

2.7.2 Métodos ou Técnicas de Mineração de Dados

Os métodos de mineração de dados são tradicionalmente divididos em aprendizado supervisionado (preditivo) e não-supervisionado (descritivo). Apesar da linha que divide essa classificação ser tênue, ou seja, alguns modelos descritivos podem ser preditivos em algum nível e vice-versa, essa distinção é útil para a compreensão dos objetivos gerais de cada método (FAYYAD; PIATETSKY-SHAPIR; SMYTH, 1996). Existem ainda variações entre esses dois tipos de aprendizado que ficaram conhecidas como aprendizado semi-supervisionado (ENGELEN; HOOS, 2020).

Aprendizagem supervisionada é basicamente sinônimo de classificação. A supervisão no aprendizado vem dos exemplos rotulados no conjunto de dados de treinamento, isto é, para eles funcionarem são necessárias duas etapas preliminares: haver uma determinada variável alvo pré-especificada e o algoritmo receber muitos exemplos nos quais o valor da variável de destino é fornecido, para que o algoritmo possa aprender quais valores da variável de destino estão associados a quais valores do variáveis preditoras. Em resumo, necessita-se de um grupo variáveis previamente rotuladas (conjunto de treinamento) para que o algoritmo possa classificar os novos dados com os rótulos previamente conhecidos (LAROSE, 2005; HAN; KAMBER; PEI, 2011).

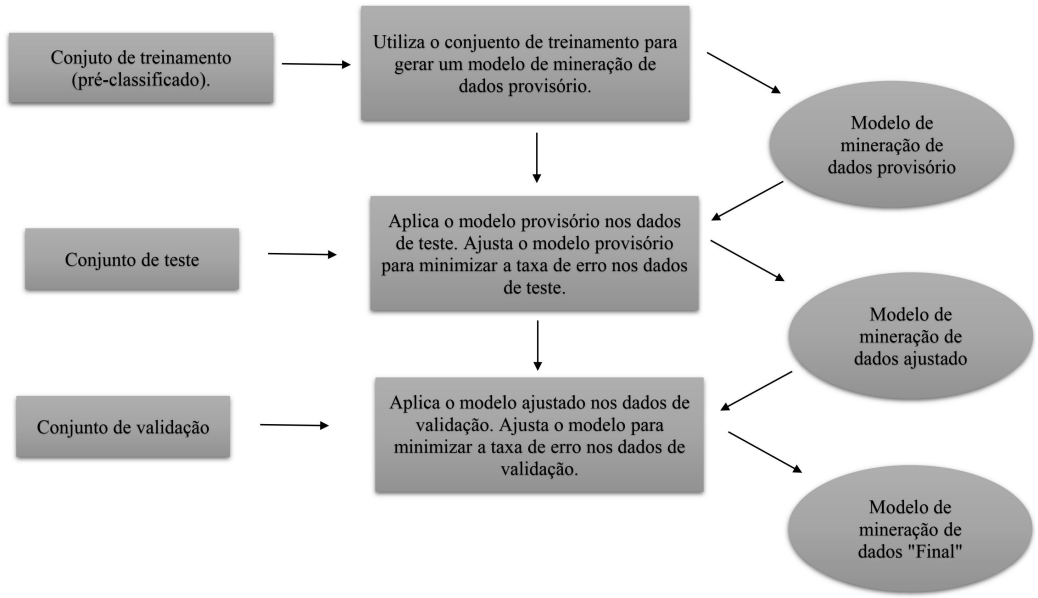

Figura 6 – Metodologia para modelos supervisionados (LAROSE, 2005).

A aprendizagem não supervisionada é essencialmente um sinônimo de agrupamento e seu objetivo é descobrir grupos similares dentro dos dados (BISHOP, 2006). O processo de aprendizagem é não-supervisionado, pois os exemplos de entrada não são rotulados por classe, isto é, não precisam de treinamento prévio para fazer classificação dos dados. Nesses métodos o algoritmo de mineração de dados procura padrões e estrutura entre to-

das as variáveis. Normalmente, utiliza-se agrupamentos para descobrir classes dentro dos dados. Por exemplo, um método de aprendizagem não supervisionada pode tomar, como entrada, um conjunto de imagens de dígitos manuscritos. Suponha que ele encontre 10 agrupamentos de dados. Esses agrupamentos podem corresponder a 10 dígitos distintos de 0 a 9, respectivamente. No entanto, uma vez que os dados de treinamento não são rotulados, o modelo aprendido não pode nos dizer o significado semântico dos agrupamentos encontrados (LAROSE, 2005; HAN; KAMBER; PEI, 2011).

O aprendizado semi-supervisionado é um ramo do aprendizado de máquina que visa combinar essas duas tarefas. Normalmente, os algoritmos de aprendizado semi-supervisionado tentam melhorar o desempenho em uma dessas duas tarefas, utilizando informações geralmente associadas à outra (ENGELEN; HOOS, 2020). Tomemos como exemplo a classificação semi-supervisionada: as instâncias rotuladas costumam ser difíceis, caras ou demoradas de se obter, pois exigem o esforço profissionais experientes. Enquanto isso, os dados não rotulados podem ser relativamente fáceis de coletar, mas há poucas maneiras de usá-los. O aprendizado semi-supervisionado resolve esse problema utilizando uma grande quantidade de dados não rotulados, junto com os dados rotulados, para construir melhores classificadores. Como o aprendizado semi-supervisionado requer menos esforço humano e oferece maior precisão, é de grande interesse tanto na teoria quanto na prática (ZHU, 2008).

A gama de métodos de mineração de dados é extensa. A seguir tem-se a descrição das principais técnicas.

2.7.2.1 Árvores de decisão (Decision Trees)

Árvore de decisão é um método de classificação composto por uma raiz (*root*), nós (*nodes*) e folhas (*leaves*), em que cada nó representa um questionamento a respeito do registo e apresenta duas ou mais respostas (caminhos), que levam a outro nó ou a uma folha, na qual o processo é finalizado e o registro classificado. A cada folha é designada uma classe que melhor representa a variável alvo. Com a árvore montada, para classificar um novo registro basta seguir o fluxo da árvore a partir da raiz até uma folha (ROKACH; MAIMON, 2005; KINGSFORD; SALZBERG, 2008).

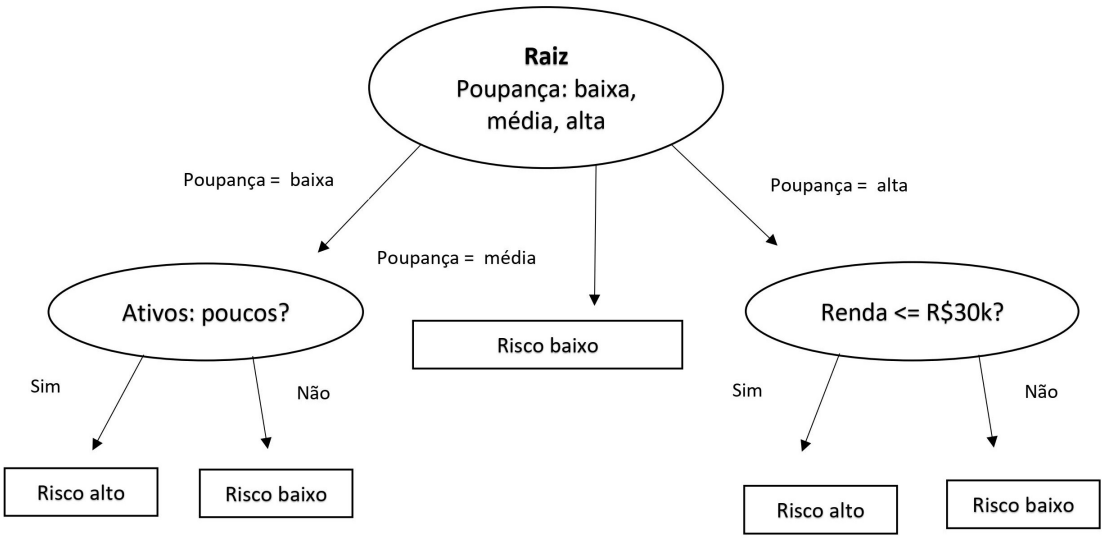

Figura 7 – Árvore de decisão simples (LAROSE, 2005).

2.7.2.2 Mineração de Padrões Frequentes (Frequent Pattern Mining)

Categorizado como uma técnica de associação, a mineração de padrões frequentes foi desenvolvida originalmente para a análise da cesta de mercado (*Market Basket*), cujo objetivo era encontrar semelhanças no comportamento de compras de clientes de supermercados e lojas online (BORGELT, 2012). Existem muitos tipos de padrões frequentes, incluindo conjuntos de itens frequentes (*frequente itemset*) e subsequências frequentes. Um conjunto de itens frequentes normalmente se refere a um conjunto de itens que muitas vezes aparecem juntos em um conjunto de dados, por exemplo, leite e pão, que são frequentemente comprados juntos no supermercado por muitos clientes. Uma subsequência frequente se refere por exemplo ao padrão que os clientes têm de comprar primeiro um laptop, seguido por uma câmera digital e, em seguida, um cartão de memória. Após ter os itens ou subsequências, regras de associação são geradas pela mineração desse conjunto. Para validar essas regras utiliza-se os conceitos de suporte e confiança. Confiança se refere à chance de a pessoa comprar o item 2 uma vez que já adquiriu o item 1, já o suporte se refere à qual porcentagem de todas as transações sob análise tiveram esses dois itens comprados juntos. Conhecimentos gerados dessa técnica oferecem grandes benefícios no processo de tomada de decisão (LUNA; FOURNIER-VIGER; VENTURA, 2019; MOENS; AKSEHIRLI; GOETHALS, 2013; HAN; KAMBER; PEI, 2011).

2.7.2.3 Redes Neurais Artificiais (Artificial Neural Networks)

É uma técnica inspirada no complexo sistema de aprendizado do cérebro animal que consiste em uma série de neurônios interconectados e pode ser aplicada na mineração de

dados para predição, classificação e agrupamento. As redes neurais artificiais representam uma tentativa, em um nível muito básico, de imitar o tipo de aprendizagem não linear que ocorre nas redes neuronais encontradas na natureza. Nessa técnica os dados de entrada de um neurônio vem da saída de outro neurônio ou da base de dados, caso seja a primeira camada, então são combinados por uma função de combinação como somatório por exemplo, em seguida são utilizadas como dados de entrada para a função de ativação (normalmente não linear) para, por fim, serem enviados para o próximo neurônio e assim por diante. A rede neural é composta de duas ou mais camadas, embora a maioria das redes consista em três camadas: uma camada de entrada, uma camada oculta e uma camada de saída. Embora possa haver mais de uma camada oculta, maioria das redes contém apenas uma, o que é suficiente para a maioria dos propósitos. O processo de aprendizado da rede neural utiliza dados de treinamento ajustando os pesos constantemente para reduzir o erro de previsão. O treinamento é um processo longo e pode ser necessárias muitas passagens do conjunto de dados pelos neurônios até se obter uma rede com os erros minimizados a ponto de ser aceitável para o propósito desejado (LAROSE, 2005; STAHL; JORDANOV, 2012).

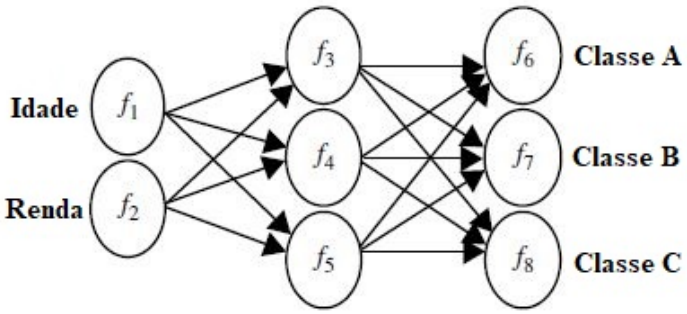

Figura 8 – Redes neurais (HAN; KAMBER; PEI, 2011).

2.7.2.4 Agrupamento Hierárquico (Hierarchical Clustering)

Nos métodos de agrupamento hierárquico, uma estrutura de grupos semelhante a uma árvore (dendrograma) é criada por meio de particionamento recursivo dos registros. Este particionamento pode ser divisivo ou aglomerativo. Os métodos de agrupamento aglomerativos inicializam com cada observação sendo um pequeno agrupamento próprio. Então, nas etapas seguintes, os dois agrupamentos mais próximos são agregados em um novo agrupamento combinado. Desta forma, o número de agrupamentos no conjunto de dados é reduzido por um em cada etapa sucessivamente até se obter a estrutura de agrupamento desejada. Métodos divisivos de agrupamento começam com todos os registros em um grande agrupamento, com a os registros que apresentam maior diferença sendo divididos

recursivamente até se obter a estrutura de agrupamento desejada (GAERTLER, 2005; ROKACH; MAIMON, 2005; LAROSE, 2005).

2.7.2.5 Agrupamento K-Mean (K-Mean Clustering)

Dado um conjunto de dados o algoritmo seleciona K registros de forma aleatória para serem os agrupamentos iniciais, então para cada registro restante é calculada a distância (similaridade) entre o registro e o centro dos agrupamentos iniciais. O registo é então colocado no agrupamento de menor distância, i.e. o agrupamento com maior similaridade (LAROSE, 2005; ROKACH; MAIMON, 2005).

2.7.2.6 K-ésimo Vizinho mais Próximo (K-Nearest Neighbor)

É um algoritmo normalmente utilizado em tarefas de classificação, mas também pode ser empregado para regressão ou predição (LAROSE, 2005). Dado um número K, este método classifica um novo registro com a classe mais comum dentre os K registros mais similares presentes em sua base de dados (HAND; MANNILA; SMYTH, 2001).

2.7.2.7 Regressão Linear e Não-Linear

Estes métodos utilizam variáveis contínuas já conhecidas, chamadas de variáveis preditoras, para fazer a predição de variáveis desconhecidas, chamadas de variáveis resposta. A ideia dessas técnicas é utilizar exemplos em que tanto as variáveis preditoras quanto as reposta são conhecidas para criar um modelo capaz de prever o valor numérico de novos casos em que apenas as variáveis preditoras são conhecidas. A regressão linear prevê valores futuros que podem ser descritos como uma combinação linear dos valores conhecidos, sendo possível criar um modelo no qual o valor desejado y, é uma função linear dos valores conhecidos x (HAND; MANNILA; SMYTH, 2001). A regressão não linear por sua vez é uma extensão da regressão linear na qual a relação entre as variáveis preditoras e as resposta é não linear podendo ser modelada por uma função polinomial (SMYTH, 2006).

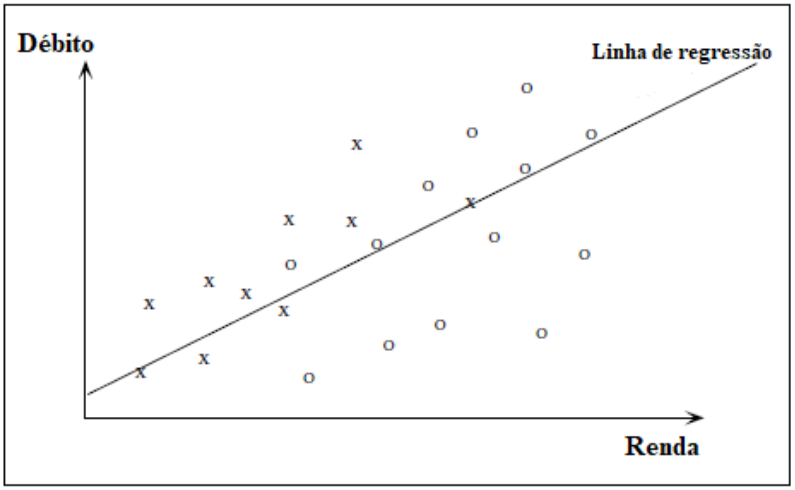

Figura 9 – Regressão linear (FAYYAD; PIATETSKY-SHAPIR; SMYTH, 1996).

2.7.2.8 Classificação Baseada em Regra (Rule-Based Classification)

É uma técnica que utiliza a estrutura "SE condição ENTÃO conclusão"para classificar registros. Essas regras podem ser geradas a partir de uma árvore de decisão ou diretamente do conjunto de treinamento utilizando algoritmos de cobertura sequencial (sequential covering). Para extrair as regras da árvore de decisão, uma regra é criada para cada caminho desde a raiz até uma folha. Tem-se um exemplo desse método na seguinte regra "SE idade = jovem E estudante = sim ENTÃO compra computador = sim"(HAN; KAMBER; PEI, 2011).

3 Metodologia

A pesquisa realizada neste trabalho utilizou dados de mão de obra, equipamentos, avanço físico e outros fatos relevantes para o processo de execução de uma obra civil, para serem analisados por ferramentas de BI. Realizou-se estas análises por meio do software Power BI, com o intuito de gerar um relatório interativo que facilitasse o acompanhamento da obra, ao mostrar as informações mais relevantes para esse processo de forma clara e intuitiva. O intuito deste relatório é permitir ao gestor averiguar, de forma rápida e fácil, os pontos que estão de acordo com o planejado e os que precisam ser melhorados ou modificados.

3.1 Os Dados

Os dados utilizados nesse trabalho foram fornecidos por uma empresa de consultoria real do setor de construção civil, que realiza o intermédio (fiscalização e gerenciamento do projeto) entre o cliente e a contratada (empreiteira). As bases de dados foram preenchidas semanalmente pela contratada de acordo com os acontecimentos da obra e com as projeções feitas por especialistas da construção civil.

Essas bases de dados contemplam as informações de maior relevância para o bom acompanhamento e monitoramento de uma obra. Mais especificamente, os dados analisados contêm informações como o avanço geral da obra, fatos relevantes, datas e marcos importantes, registros de acidentes, notas de qualidade e SSMA (Saúde, Segurança e Meio Ambiente), não conformidades, quantidades de mão de obra direta, indireta e maquinário e os avanços específicos de quatro serviços notáveis.

As bases de dados advindas da empresa contratada foram compiladas, transformadas para as bases de dados da empresa e verificadas, para então, serem utilizadas e atualizadas no software Power BI, i.e., as bases de dados devem ser muito bem estruturadas e organizadas e a atualização dos dados deve ser realizada com rigor.

3.2 Construção das Análises

O Power BI é uma ferramenta de análise de dados extremamente poderosa, que consegue expor informações pertinentes à situação em questão de forma simples, rápida e interativa, mas para isso, é necessário conhecer muito bem os dados disponíveis e o problema a ser resolvido. Portanto, essa seção do trabalho é destinada a demonstrar o que foi feito para se obter os resultados do capítulo seguinte. Dessa forma, para realizar uma análise que possibilitasse a obtenção de informações relevantes que pudessem auxiliar os responsáveis pela obra a gerenciá-la da melhor maneira, foi necessário cumprir três etapas: entender o problema, preparar as bases de dados e modelar o problema no Power BI.

Sob posse dos dados, o desafio seguinte consiste em estudar o material a fim de ter um conhecimento mais aprofundado das informações que podem ser obtidas com os dados disponíveis. Nessa etapa procurou-se definir quais perguntas devem ser respondidas e o que se tinha disponível para respondê-las.

A etapa seguinte consiste em preparar as bases de dados para um modelo aceito pelo Power BI. Nessa etapa, os dados recebidos em Excel foram tratados, de forma a verificar inconsistências e adequando-os para os modelos aceitos pelo Power BI. Para tal, viu-se a necessidade de criar tabelas auxiliares e modificar as existentes.

A terceira etapa consiste em modelar o desafio proposto no Power BI. Pra isso utilizou-se as ferramentas de análise de dados descritiva disponíveis no programa como gráficos, histogramas, tabelas, matrizes, medidas de síntese e índices.

Tendo passado por essas etapas, visando alcançar o objetivo proposto, viu-se a necessidade de criar um relatório interativo de 7 páginas abrangendo todas as informações pertinentes disponíveis nos dados coletados. O relatório encontra-se dividido da seguinte forma:

Avanço Geral: contém informações do andamento da obra como um todo.

Serviços Notáveis: quatro relatórios apresentando os quatro serviços mais importantes da obra.

Histograma de Recursos: apresenta a evolução dos recursos de mão de obra e equipamentos durante a obra.

Fatos Relevantes e Qualidade: apresenta os fatos mais relevantes ocorridos na obra, bem como avaliações dos setores de qualidade, SSMA (Saúde Segurança e Meio Ambiente), não conformidades e acidentes.

4 Apresentação e discussão dos resultados

Dada a metodologia apresentada na seção anterior, criou-se um relatório que permite aos gestores acompanhar a obra, tendo as informações mais relevantes dispostas de maneira simples, interativa e atualizadas semanalmente. As figuras a seguir apresentam o relatório interativo construído no software Power BI.

4.1 Menu Inicial

A Figura 10 apresenta o "Menu" inicial, em que são apresentados as 7 páginas do relatório, cuja navegação é simples e intuitiva. Para acessar a página desejada, basta clicar no ícone correspondente.

Figura 10 – Menu do relatório.

4.2 Avanço Físico - Geral

A Figura 11 apresenta a primeira página do relatório, denominada "Avanço Físico - Geral", que contém os dados de avanço físico geral da obra. Nesse painel, é possível acompanhar o avanço físico da obra por meio de um gráfico do tipo Curva S, também conhecido como Curva Logística.

Muito utilizada no planejamento, programação e controle de projetos, a Curva S mostra o comportamento da distribuição de um recurso ou população de forma cumulativa.

A curva apresenta o projeto como um todo em termos dos avanços físicos necessários para a conclusão da obra. Com ela, é possível verificar o que foi planejado inicialmente, bem como o que foi de fato realizado e, com isso, reprogramar os avanços que necessitam ser realizados nas semanas seguintes para se concluir a obra no prazo estipulado.

Além da Curva S, é possível filtrar as informações da página por semana, para se ter um visão mais detalhada do andamento da obra em um período específico. O detalhamento de informações mostra dados como o desvio semanal e acumulado e a aderência da curva "Real Acumulado" em relação ao "Planejado Acumulado". O desvio por sua vez, revela a diferença entre o planejado e o realizado em termos percentuais. Todos esses dados mudam de cor de acordo com o nível de satisfação que aquele número reflete: a cor verde é utilizada caso o dado evidencie um desempenho favorável, o amarelo, mediano e o vermelho, ruim.

A análise desse painel revela a presença de um desvio negativo considerável, a partir da Semana 8. Esse fato sinaliza para o gestor que os avanços não estão ocorrendo conforme o planejamento inicial, permitindo que ele tome decisões em tempo hábil para reverter essa situação desfavorável durante as próximas semanas. Com as reprogramações realizadas, percebe-se que o desvio futuro tende a zero próximo ao fim da obra, não afetando o término previsto.

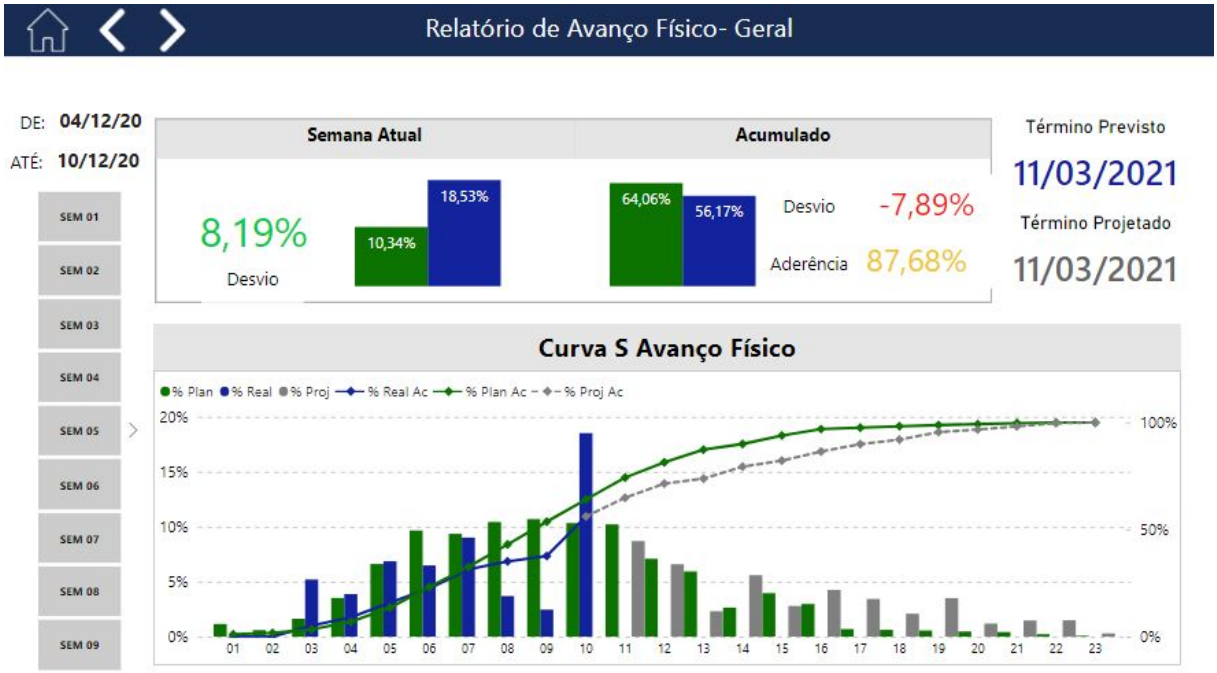

Figura 11 – Relatório de Avanço Físico - Geral.

4.3 Serviços Notáveis

As Figuras 12, 13, 14 e 15 apresentam os 4 serviços mais importantes da obra, denominados "Serviços Notáveis". As informações dessas páginas do relatório seguem a mesma ideia do painel de "Avanço Físico - Geral", em que os avanços físicos semanais, em m^3, necessários para a conclusão da obra no prazo planejado, estão dispostos no tempo. Nesse painel observa-se, além da Curva S, as principais informações relacionadas ao serviço em questão, dispostas no formato de cartões no lado direito da página. Esses cartões contemplam dados do total de m^3 planejados para a obra, do avanço realizado até a data atual, quanto o avanço físico realizado representa do total, qual o avanço planejado para a semana em questão, qual o avanço realizado na semana em questão, o avanço total planejado até a semana em questão, o avanço total realizado até a semana em questão e a aderência da curva dos avanços realizados em relação ao que foi planejado inicialmente. As 5 últimas informações supracitadas podem ser filtradas por semana, para se ter uma visão mais detalhada dos acontecimentos relacionados ao serviço que se está analisando.

A Figura 12 demonstra um serviço em que a curva do "Avanço Real" e do "Avanço Projetado" estão seguindo o que foi planejado inicialmente, não necessitando de intervenções em relação ao projeto inicial.

Já as Figuras 13, 14 e 15 mostram serviços cuja curva do "Avanço Projetado" tende a se descolar do planejado, havendo a necessidade da elaboração de um plano de ação, a fim de evitar atrasos e prejuízos no futuro. Sendo que Figuras 13 e 15 ilustram situações que mesmo tendo sido observado desvios futuros, até o período analisado, o término projetado para o fim da obra mostra-se atrasado em duas semanas. Esses são pontos de atenção que precisam ser trabalhados pelo gestor caso seja imprescindível cumprir os prazos estabelecidos inicialmente.

Figura 12 – Serviço Notável 1.

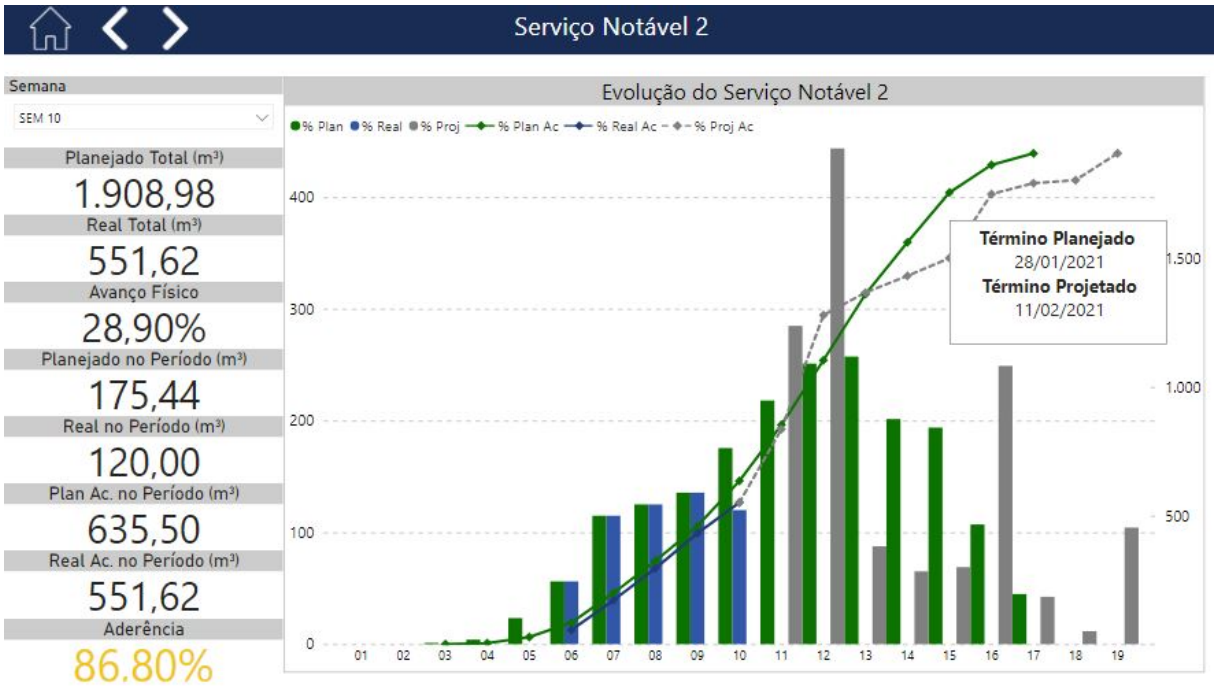

Figura 13 – Serviço Notável 2.

Figura 14 – Serviço Notável 3.

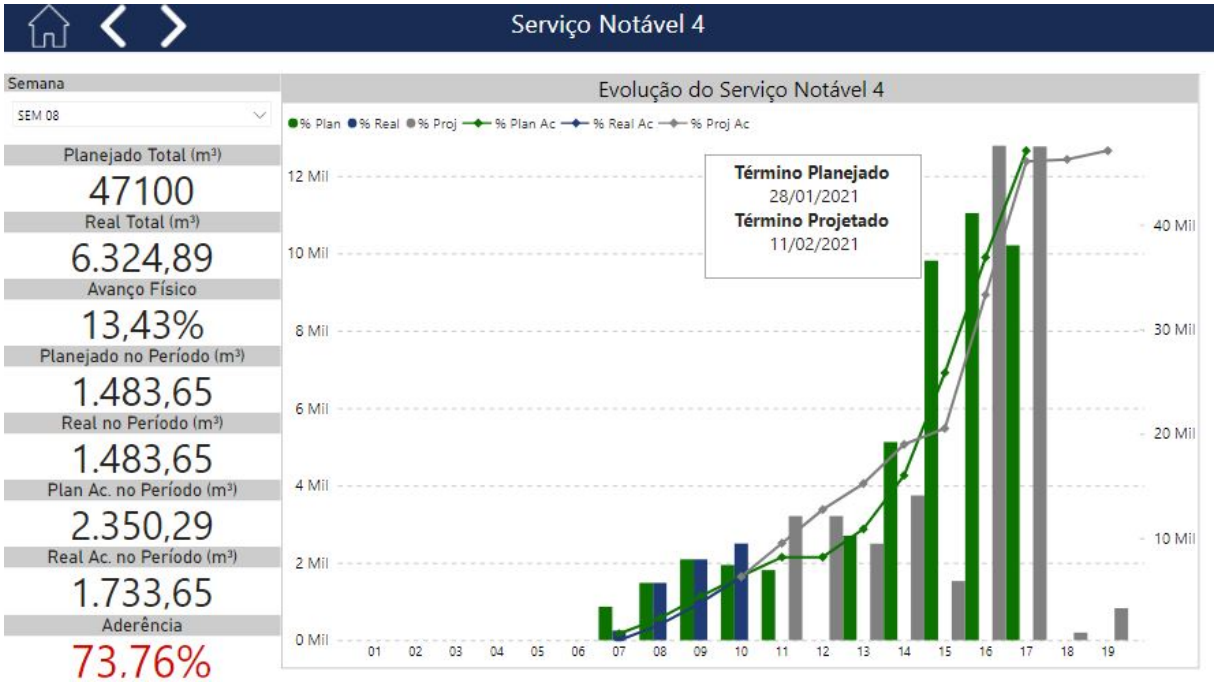

Figura 15 – Serviço Notável 4.

4.4 Histogramas de Recursos

Outro dado considerado relevante para um bom gerenciamento de uma obra civil neste trabalho, se trata da evolução dos recursos ao longo dos meses.

A sexta página do relatório (Figura 16) apresenta três histogramas de recursos, que contemplam a evolução dos recursos de mão de obra direta, mão de obra indireta e equipamentos necessário para a execução da obra.

Cada histograma apresenta a quantidade planejada do recurso em questão, a quantidade que de fato foi utilizada até a data da análise e a projeção do que ainda será necessário, de acordo com as eventualidades ocorridas ao longo da execução do projeto. Observa-se também, nos cartões de aderência no canto esquerdo, a relação da utilização real do recurso com o planejado. Essas informações podem ser filtradas por semana e apresentam os dados com cores distintas, de acordo com o nível de satisfação que aquele número reflete: a cor verde é utilizada caso o dado evidencie um desempenho favorável, o amarelo, mediano e o vermelho, ruim.

A análise desse painel mostra que a evolução dos recursos está seguindo de acordo com o planejado, apesar de haver pequenas variações. Sinalizando um ponto que deve estar no radar do gestor para evitar maiores desvios no futuro.

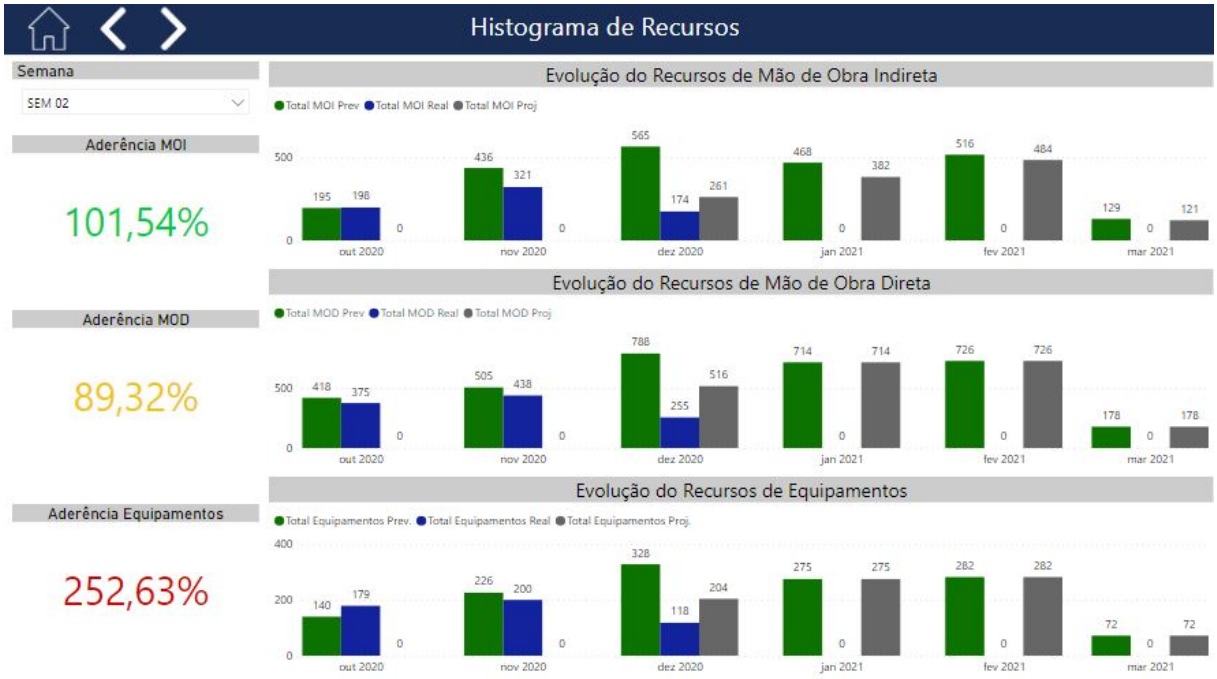

Figura 16 – Histograma de Recursos.

4.5 Fatos Relevantes, SSMA & Qualidade , Acidentes e Não Conformidades

A última página do relatório (Figura 17) apresenta fatos relevantes que aconteceram no decorrer da obra, as notas avaliativas dos setores de SSMA (Saúde, Segurança e Meio Ambiente) e de qualidade, um gráfico com as não conformidades e um diagrama de acidentes.

Os “Fatos Relevantes” englobam acontecimento de importância considerável ocorridos no desenrolar da obra, como a execução de uma atividade importante e a realocação de recursos entre as frentes de trabalho. Esses fatos são classificados nas categorias de ”Civil” ou ”Geral” e estão dispostos de acordo com a data do ocorrido.

Os gráficos com as notas avaliativas dos setores apresentam a meta esperada para cada setor e a nota de fato recebida. A análise desses dados permite perceber uma queda nas notas de Saúde e Segurança, Laboratório e Meio Ambiente, revelando pontos que necessitam de uma maior atenção, ou até mesmo um replanejamento, para evitar problemas futuros no desempenho desses setores.

O Gráfico de Rosca contém as informações dos registos de Não Conformidades, que são desvios que ocorrem na execução de qualquer processo, não atendendo o padrão de qualidade pré-estabelecido. De forma muito visual, este gráfico apresenta o total de Não Conformidades registradas na obra, bem como o percentual dos registos que estão em aberto e os que já foram concluídos.

As últimas informações do relatório estão relacionadas ao número de acidentes ocorridos na obra. Este diagrama apresenta os números de “Quase Acidentes”, acidentes categorizados como “SPT” (Sem Perda de Tempo), em que o acidentado sofre lesão que não o impeça de voltar ao trabalho até o dia seguinte e “CPT” (Com Perda de Tempo), em que o acidentado sofre lesão que o impede de voltar ao trabalho no dia imediato ao do acidente. Essas são informações de fundamental importância, uma vez que possibilitam identificar se está havendo uma quantidade anormal de acidentes, para que possam ser investigados os serviços e operações que apresentam os maiores riscos para os funcionários, com a intenção de modificá-los, visando aumentar a segurança nessas atividades, evitando futuros acidentes.

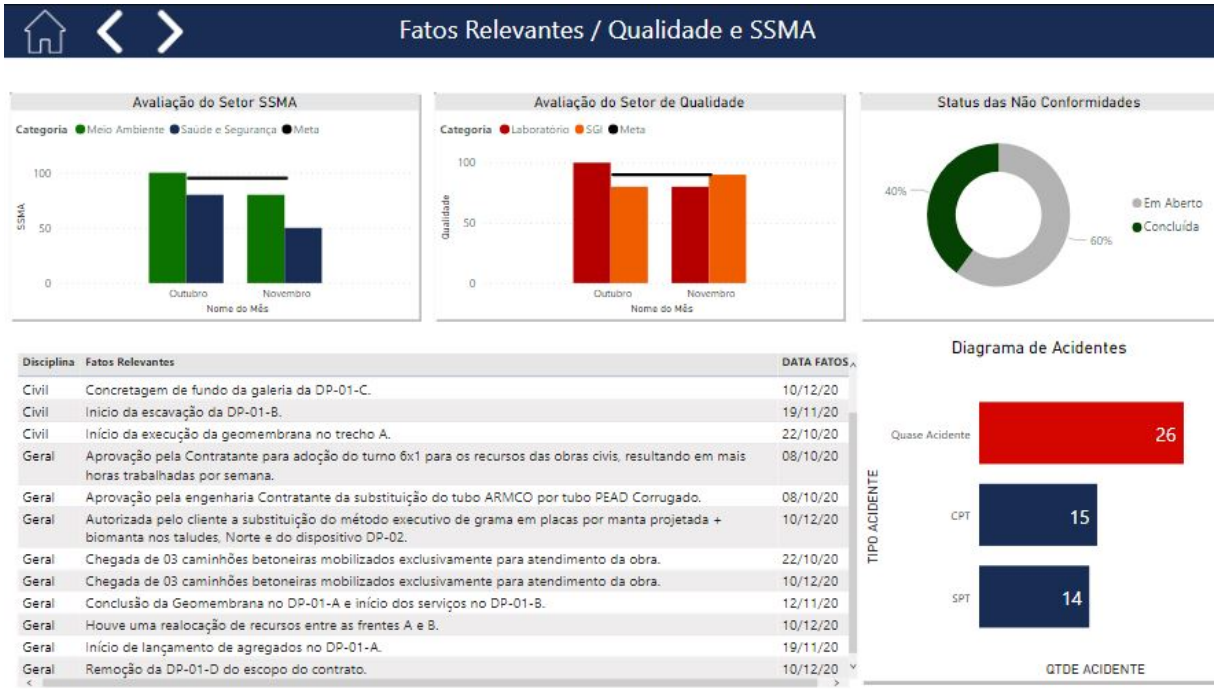

Figura 17 – Fatos Revelantes / Qualidade & SSMA / Acidentes / Não Conformidades.

5 Conclusões e considerações finais

A literatura estudada para o desenvolvimento deste trabalho expõe diversas melhorias e benefícios alcançados pelas empresas a partir da aplicação de técnicas e programas de Inteligência de Negócios em suas atividades, dentre as quais cabe citar: o aprimoramento da tomada de decisão; oportunismo; melhor qualidade de informação; economia de tempo e de custos; prevenção de custos; aumento de receita; aumento no índice de acertos; e maior precisão e capacidade de ação gerada pelas informações produzidas pelo BI.

A utilização do programa Power BI permitiu a realização de análises descritivas dos dados coletados de uma obra civil. Sendo que os resultados obtidos nesse trabalho ratificam a literatura existente sobre o tema, uma vez que, o relatório desenvolvido nesse estudo permite observar que com a análise descritiva dos dados de uma obra tem-se os seguintes benefícios:

Velocidade na Geração do Relatório: ter todos os dados consolidados em um único lugar, confere celeridade ao processo de criação do relatório para apresentação dos dados referentes aos avanços e acontecimentos da semana para o cliente. O que antes era realizado com Excel e Project e apresentado em Power Point, agora fica unificado no Power BI.

Acelera a Tomada de Decisão: ter os principais dados relacionados à obra expostos de maneira simples, visual e interativa facilita o entendimento e a visualização do andamento da obra como um todo e dos pontos que precisam de uma maior atenção para evitar desvios no planejamento inicial do projeto.

Agrega Valor ao Serviço Prestado ao Cliente: a apresentação de um relatório interativo com gráficos, histogramas, tabelas e cartões que mudam de cor de acordo com o desempenho verificado, confere um maior valor ao serviço prestado ao cliente, do que apresentar apenas um relatório com números e textos descritivos de entendimento mais difícil.

Aprimora o Processo de Gestão do Projeto: o alto poder de processamento de dados e a capacidade de armazenar e monitorar o histórico ao longo do tempo aperfeiçoa o processo de gerenciamento do projeto.

Aumenta o Índice de Acerto: ter a informação certa no momento certo aumenta o índice de acertos das tomadas de decisão.

Este trabalho mostrou alguns benefícios que a utilização de ferramentas de Inteligência de Negócios pode trazer para o processo de gerenciamento de uma obra civil. Seria

interessante observar em estudos futuros como as análises preditiva e prescritiva poderiam ser aplicadas no setor da construção civil e quais os benefícios obtidos com elas.

Referências

ALASPURKAR, M. A. K. and. S. J. Data mining technique to analyse the metrological data. *International Journal of Advanced Research in Computer Science and Software Engineering*, v. 3, n. 2, p. 114–118, 2013.

ARGOTTE, L.; MEJIA-LAVALLE, M.; SOSA, R. Business intelligence and energy markets: A survey. *Conference: Intelligent System Applications to Power Systems*, v. 15, 2009.

BISHOP, C. M. *Pattern recognition and machine learning*. Berlim: Springer, 2006.

BORGELT, C. Frequent item set mining. *WIREs Data Mining and Knowledge Discovery*, v. 2, n. 6, p. 437–456, 2012.

DAVISON, L. Measuring competitive intelligence effectiveness: Insights from the advertising industry. *Competitive Intelligence Review*, v. 2, n. 4, p. 25–38, 2001.

ENGELEN, J. E. van; HOOS, H. H. A survey on semi-supervised learning. *Machine Learning*, v. 109, n. 2, p. 373–440, 2020. Disponível em: <https://doi.org/10.1007/s10994-019-05855-6>.

FAYYAD, U.; PIATETSKY-SHAPIR, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, n. 3, p. 37–54, 1996.

GAERTLER, M. Network analysis. *Lectures Notes in Computer Science*, p. 127–215, 2005.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques. 3rd Edition*. Waltham: Morgan Kaufmann, 2011.

HAND, D.; MANNILA, H.; SMYTH, P. *Principles of Data Mining*. Cambridge - MA: Bradford Book, 2001.

HANNULA, M.; PIRTTIMäKI, V. Business intelligence: Empirical study ont the top 50 finnish companies. *Journal of American Academy of Business*, Cambridge, v. 2, n. 2, p. 593, 2003.

HARISON, E. Critical success factors of business intelligence system implementations: Evidence from the energy sector. *International Journal of Enterprise Information Systems*, v. 8, n. 2, p. 1–13, 2012.

ISIK Öykü; JONES, M. C.; SIDOROVA, A. Business intelligence success: The roles of bi capabilities and decision environments. *Information and Management*, v. 50, p. 13–23, 2013.

KHAN, R.; QUADRI, S. M. K. Business intelligence: An integrated aproach. *Business Intelligence Journal*, v. 5, n. 1, p. 64–70, 2012.

KINGSFORD, C.; SALZBERG, S. L. What are decision trees? *Nature Bioechnology*, v. 26, n. 9, p. 1011–1013, 2008.

LAROSE, D. T. *Discovering Knowledge in Data: An Introduction to Data Mining*. Nova Jersey: John Wiley and Sons, 2005.

LöNNQVIST, A.; PIRTTIMäKI, V. The measurement of business intelligence. *Information Systems Management*, v. 23, n. 1, p. 32, 2006.

LUNA, J. M.; FOURNIER-VIGER, P.; VENTURA, S. Frequent itemset mining: A 25 years review. *WIREs Data Mining and Knowledge Discovery*, v. 9, n. 6, p. 1–15, 2019.

MOENS, S.; AKSEHIRLI, E.; GOETHALS, B. Frequent itemset mining for big data. *IEEE International Conference on Big Data*, p. 111–118, 2013.

PIRTTIMäKI, V.; LöNNQVIST, A.; KARJALUOTO, A. J. Measurement of business intelligence in a finnish telecommunications company. *The Electronic Journal of Knowledge Management*, v. 4, n. 1, p. 83–90, 2006.

REIS, E. A.; REIS, I. A. Análise descritiva de dados. 2002. Disponível em: <http://www.est.ufmg.br/portal/arquivos/rts/rte0202.pdf>.

ROKACH, L.; MAIMON, O. *Data Mining and Knowledge Discovery Handbook*. Boston - MA: Springer, 2005.

SHARDA, R.; DELEN, D.; TURBAN, E. *Business Intelligence e Análise de Dados para Gestão do Negócio. 4. ed.* Porto Alegre: Bookman, 2019.

SMYTH, G. K. Nonlinear regression. *Encyclopedia of Envirometrics*, p. 201–213, 2006.

STAHL, F.; JORDANOV, I. An overview of the use of neural networks for data mining tasks. *WIREs Data Mining and Knowledge Discovery*, v. 2, n. 6, p. 193?208, 2012.

ZHU, X. Semi-supervised learning literature survey. *Computer Sciences TR 1530*, University of Wisconsin ? Madison, 2008.