

PONTÍFICA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Beatriz Perla Benezra Dehtear

**MÉTODO PARA O DESENVOLVIMENTO DA INTELIGÊNCIA ORGANIZACIONAL EM
ORGANIZAÇÕES ORIENTADAS A DADOS**

Porto Alegre
2019

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica
do Rio Grande do Sul

PONTÍFICA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**MÉTODO PARA O DESENVOLVIMENTO DA INTELIGÊNCIA
ORGANIZACIONAL EM ORGANIZAÇÕES ORIENTADAS A
DADOS**

Beatriz Perla Benezra Dehtear

Dissertação apresentada como
requisito parcial à obtenção do
grau de Mestre em Ciência da
Computação na Pontifícia
Universidade Católica do Rio
Grande do Sul.

Orientador: Prof. Dr. Duncan Dubugras Alcoba Ruiz

Porto Alegre

2019

Ficha Catalográfica

D323m Dehtear, Beatriz Perla Benezra

Método para Desenvolvimento da Inteligência Organizacional em Organizações Orientadas a Dados / Beatriz Perla Benezra Dehtear . – 2019. 177 f.

Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientador: Prof. Dr. Duncan Dubugras Alcoba Ruiz.

1. Organizações Orientadas a Dados. 2. Descoberta de Conhecimento em Banco de Dados. 3. Organizações Inteligentes. 4. Cultura Orientada a Dados. I. Ruiz, Duncan Dubugras Alcoba. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS com os dados fornecidos pelo(a) autor(a).

Bibliotecária responsável: Salete Maria Sartori CRB-10/1363

Beatriz Perla Benezra Dehtear

**MÉTODO PARA O DESENVOLVIMENTO DA INTELIGÊNCIA
ORGANIZACIONAL EM ORGANIZAÇÕES ORIENTADAS A DADOS**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação, Escola Politécnica da Pontifícia Universidade Católica do Rio Grande do Sul.

Aprovado em _____ de _____ de 20____.

BANCA EXAMINADORA:

Profa. Dra. Sabrina Marczak (PUCRS)

Profa. Dra. Edimara Mezzomo Luciano (PUCRS)

Prof. Dr. Duncan Dubugras Alcoba Ruiz (PPGCC/PUCRS - Orientador)

“Quando acreditamos apaixonadamente em algo que ainda não existe, nós o criamos. O inexistente é o que não desejamos o suficiente”

Franz Kafka (1883 – 1924)

AGRADECIMENTOS

Agradeço a minhas filhas Gabriela, Micaela e Fernanda pela compreensão e pelo apoio incondicional. Elas são a minha constante fonte de energia e inspiração! Em especial agradeço a Micaela, pelo seu apoio profissional como Designer na elaboração do Tutorial.

Agradeço ao meu orientador, Prof. Duncan, pelo seu voto de confiança, pelo conhecimento adquirido com a sua constante e valiosa orientação e pela sua infinita paciência.

Agradeço também às professoras integrantes da mesa, Profa. Sabrina e Profa. Edimara, pelas suas avaliações e suas muito valiosas observações.

Agradeço a todos os professores que durante as aulas me brindaram com seu infinito conhecimento e grande experiência. Cada um deles fez uma grande diferença!

Agradeço a todos os profissionais que tão gentilmente dedicaram seu tempo para participar da avaliação deste trabalho.

Por fim, a todas aqueles amigos e colegas para os quais faltei e souberam me entender a apoiar nesta realização, meu muito obrigada!

MÉTODO PARA O DESENVOLVIMENTO DA INTELIGÊNCIA ORGANIZACIONAL EM ORGANIZAÇÕES ORIENTADAS A DADOS

RESUMO

Embora a Cultura Orientada a Dados seja amplamente adotada pelas organizações como uma forma de apoio à tomada de decisão, existe um potencial maior nos dados que pode ser explorado na geração de *insights* e novos conhecimentos. Este novo conhecimento, quando aplicado em prol dos objetivos organizacionais, imprime um comportamento organizacional diferenciado que caracteriza as chamadas “Organizações com Comportamento Inteligente”. Existe uma relação direta entre a complexidade dos diferentes elementos envolvidos para a extração de conhecimento a partir de dados e o valor que o conhecimento gerado tem para a organização. Assim sendo, existe um processo evolutivo de complexidade crescente na passagem da situação inicial, onde os dados são usados no apoio ao processo de tomada de decisão, para uma nova situação, onde estes dados são usados para descoberta de conhecimento. Este trabalho apresenta um Método que tem como objetivo ajudar organizações a fazer a transição entre as duas situações descritas. O Método considera como ponto de partida uma plataforma DW/BI (*Data Warehouse/Business Intelligence*). Um conjunto de ações realizadas em forma organizada executam a transição para o estado final, resultando na implementação do processo de KDD (*Knowledge Discovery in Databases*). Na transição, a operação da plataforma DW/BI não é afetada. O Método é avaliado pela sua aplicabilidade e facilidade de uso de acordo com o questionário que faz parte do modelo TAM (*Technology Acceptance Model*) e que tem este objetivo. Um Tutorial é usado como artefato para a avaliação, esta realizada por um Grupo Focal de especialistas.

Palavras-chave: Organizações Orientadas a Dados, Descoberta de Conhecimento em Banco de Dados, Organizações Inteligentes, Cultura Orientada a Dados.

A METHOD FOR ORGANIZATIONAL INTELLIGENCE DEVELOPMENT IN DATA-DRIVEN ORGANIZATIONS

ABSTRACT

Although Data-driven culture is widely adopted by organizations as a support in the decision-making process, there is a greater potential in data that can be explored by the generation of insights and new knowledge. This new knowledge, when applied in favor of organizational objectives, creates a differentiated organizational behavior that characterizes the so called "Intelligent Behavior Organizations". There is a direct relationship between the complexity of the different elements involved in the extraction of knowledge from data and the value that the knowledge generates for the organization. Thus, there is an evolutionary process of increasing complexity in the transition of the situation, in which data is used to support the decision-making process, into a new situation in which data is a knowledge discovery resource. This work presents a Method that aims helping organizations with the transition between the two described situations. The Method considers as the starting point a DW/BI (Data Warehouse/Business Intelligence) platform. A set of actions performed in an organized way executes the transition to the final state resulting in the implementation of the Knowledge Discovery in Databases (KDD) process. During the transition, the operation of the DW/BI platform is unaffected. The Method is evaluated considering its applicability and its ease of use according to TAM's (Technology Acceptance Model) questionnaire, that is part of this model. A Tutorial is used as an artifact for evaluation by a specialists' Focal Group.

Keywords: Data Oriented Organizations, Knowledge Discovery in Databases, Smart Organizations, Data-driven culture.

LISTA DE FIGURAS

Figura 1 - Modelo de BI, Kimball [KIM13]	29
Figura 2 - Modelos de KDD de acordo com a pesquisa de Mariscal [MAR10].....	31
Figura 3 - Processo de KDD, Fayyad et al. [FAY96].....	32
Figura 4 - Modelo CRISP-DM [CHA00].....	34
Figura 5 - Diagrama evolutivo do modelo de Kimball [KIM13] realizado pelo autor	37
Figura 6 - Estrutura do Método proposto.....	42

LISTA DE TABELAS

Tabela 1 - Tabela comparativa dos modelos de Inmon e Kimball com foco na metodologia e Arquitetura, BRESLIN [BRE04]	26
Tabela 2 - Tabela comparativa dos modelos de Inmon e Kimball com foco na natureza do processo de tomada de decisão, BRESLIN [BRE04]	27
Tabela 3 - Análise dos elementos dos estados inicial e final na composição do Método	39
Tabela 4 - Conhecimento necessário para o Usuário do Método.....	45
Tabela 5 - Ações que implementam o Método proposto	47
Tabela 6 - Critérios de Utilidade Propostos pelo Modelo TAM [DAV89]	54
Tabela 7 - Critérios de Facilidade de uso Propostos pela Modelo TAM [DAV89]	55
Tabela 8 - Adaptações realizadas pelo Autor na Tabela de Utilidade do Modelo TAM	56
Tabela 9 - Adaptações realizadas pelo Autor na Tabela de Facilidade de Uso do Modelo TAM	57
Tabela 10 - Avaliação dos Domínios de Conhecimento dos participantes do Grupo Focal	61
Tabela 11 - Resultado da avaliação da Utilidade do Método aplicando TAM.....	62
Tabela 12 - Resultado da avaliação da Facilidade de Uso do Método aplicando TAM	62

LISTA DE ABREVIATURAS E SIGLAS

BI – *Business Intelligence*

DW/BI - *Data Warehouse/Business Intelligence*

CRM – *Customer Relationship Management*

DMM – *Data Maturity Model*

DW – *Data Warehouse*

ETL - *Extract, Transform and Load*

KDD – *Knowledge Discovery in Databases*

OLAP - *On-line Analytical Processing*

SAD – *Sistema de Apoio à Decisão*

TAM - *Technology Acceptance Model*

SUMÁRIO

1. INTRODUÇÃO	13
1.1. MOTIVAÇÃO	17
1.2. DESENHO DA PESQUISA	18
1.3. ORGANIZAÇÃO DO TRABALHO	19
2. FUNDAMENTAÇÃO TEÓRICA	20
2.1. O CONCEITO DE INTELIGÊNCIA	20
2.2. INTELIGÊNCIA INDIVIDUAL	20
2.3. INTELIGÊNCIA COLETIVA E ORGANIZACIONAL	22
2.4. DA INFORMAÇÃO AO CONHECIMENTO.....	23
2.5. BI E SEUS MODELOS	25
2.5.1. O Modelo de Kimball	27
2.6. O PROCESSO DE KDD	29
2.7. MODELOS DE KDD	30
2.7.1. O Modelo de Fayyad	32
2.7.2. O Modelo CRISP-DM	33
2.8. CONSIDERAÇÕES	35
3. MÉTODO PARA DESENVOLVIMENTO DA INTELIGÊNCIA ORGANIZACIONAL EM ORGANIZAÇÕES ORIENTADAS A DADOS.....	36
3.1. ESTRUTURA DO MÉTODO	36
3.2. DESENVOLVIMENTO DO MÉTODO	40
3.2.1. O contexto de aplicação do Método	40
3.2.2. Estrutura do Método	40
3.2.3. Impactos da aplicação do Método	41
3.2.4. Perfil do Usuário	44
3.2.5. Ações que implementam o Método	46
3.3. CONCLUSÕES	52
4. AVALIAÇÃO DO MÉTODO	53
4.1. MÉTODO DE AVALIAÇÃO DE TAM	53
4.2. GRUPOS FOCAIS	56

4.3. O TUTORIAL	57
4.4. EXECUÇÃO DA AVALIAÇÃO	57
4.4.1. Avaliação Inicial	57
4.4.2. Avaliação do Grupo Focal.....	58
4.4.2.1. Preparação	59
4.4.2.2. Execução	60
4.4.2.3. Resultados Obtidos	61
4.4.2.4. Conclusões	64
4.5. AÇÕES	65
5. EXERCÍCIO DE APLICAÇÃO DO MÉTODO	66
5.1. CONSIDERAÇÕES SOBRE A APLICAÇÃO DO MÉTODO	66
6. TRABALHOS RELACIONADOS	70
7. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	72
REFERÊNCIAS.....	74
APÊNDICE A - O TUTORIAL	78
APÊNDICE B - EXERCÍCIO DE APLICAÇÃO DO MÉTODO	151
APÊNDICE C – QUESTIONÁRIOS PARA AVALIAR O MÉTODO	176

1. INTRODUÇÃO

De acordo com Virgolim [VIR14], a inteligência ou o comportamento inteligente possui uma série de atributos, quais sejam:

- É uma propriedade que caracteriza o comportamento de um indivíduo dentro do seu contexto de atuação.
- Depende da capacidade de adaptação e das respostas do indivíduo em novos contextos.
- Está relacionada com a capacidade do indivíduo em ter sucesso em relação a um objetivo.
- Está totalmente vinculado à capacidade de aprender do indivíduo.

A inteligência é normalmente percebida como uma capacidade individual. Porém, pensando em atividades em que a união de indivíduos é necessária para a execução bem-sucedida de uma tarefa, como por exemplo construir um software, fica complicado pensar na inteligência em caráter individual, tornando-se evidente seu caráter coletivo.

Wilensky [WIL15] reúne os principais elementos e descobertas de pesquisas sobre organizações e em meados de 1960 conclui que o conhecimento determina parcialmente as decisões organizacionais e seus resultados. Ele chama esse conhecimento de "inteligência", e acrescenta que a inteligência organizacional se traduz na capacidade de processar, interpretar e comunicar a informação.

McMaster [MCM16] define a “Organização Inteligente” como a organização que tem a capacidade de reunir, gerar e disponibilizar informação para construir o conhecimento que lhe fornece subsídios para um comportamento diferenciado. Choo [CHO06] reforça o vínculo entre informação e comportamento inteligente das organizações afirmando que a informação é um componente intrínseco de quase tudo o que a organização faz. A clara percepção dos processos que transformam informação em conhecimento e posteriormente em ação faz com que as organizações sejam capazes de perceber a importância de suas fontes tecnológicas de informação.

O Instituto MIT Sloan [KIR12] define “Organizações com Cultura Orientada a Dados” como organizações que compartilham a crença de que a criação e a

aplicação prática de determinado tipo de dados e informação jogam um papel crítico em seu sucesso. Ainda outro estudo do MIT Sloan em parceria com a SAS [KIR13], mostra que o uso de modelos de Descoberta de Conhecimento em Banco de Dados (*Knowledge Discovery Databases*), geram resultados diferenciados nos processos organizacionais onde o novo conhecimento é aplicado. Cao [CAO17] aponta que em diversos tipos de organizações, algumas atividades que se apoiam em KDD, comprovadamente, apresentam resultados diferenciados. Alguns exemplos apresentam-se logo a seguir:

- Pontuação de crédito, para estabelecer a capacidade de crédito de um cliente que solicita um empréstimo.
- Detecção de fraude, para identificar transações fraudulentas e comportamento suspeito.
- Cuidados de saúde, para detectar serviços excessivos, falta de serviço, fraude e eventos como epidemias.
- Seguro, para detectar reclamações fraudulentas e avaliar riscos.
- Análise do processo de fabricação, para identificar as causas dos problemas de fabricação e otimizar os processos.
- Marketing e vendas, para identificar potenciais clientes e estabelecer a eficácia das campanhas.
- Negociação de portfólio, para otimizar uma carteira de instrumentos financeiros, maximizando os retornos e minimizando os riscos.
- Vigilância, para detectar intrusões, alterações, e conexões de dados em sensoriamento remoto.

Em 2017 o MIT Sloan [RAN18] realizou um estudo no qual entrevistou 1919 gestores pertencentes a organizações com Cultura Orientada a Dados. Esta pesquisa concluiu que 59% dos entrevistados atribuem vantagem competitiva significativa à sua maturidade, em relação ao uso de dados. O estudo acrescenta que este percentual teve um aumento significativo se comparado com a pesquisa realizada em 2015, onde responderam da mesma forma 51% dos entrevistados. O mesmo estudo afirma que esse movimento ascendente é o resultado de muitos fatores, entre os quais se destaca o fato de que o uso de descoberta de

conhecimento em dados permite entrar em novos mercados e que para manter-se nestes novos mercados é necessário incrementar esta maturidade. O autor acrescenta que entende-se como “maturidade em uso de dados” a capacidade de conseguir um melhor aproveitamento dos dados na extração de conhecimento, em prol dos objetivos organizacionais. Este conceito é aplicado por diversos autores citados neste trabalho como, Robert [ROB11], CGUAN [CHU10], Russom [RUS11], e Carnegie Mellon [CMM14].

De acordo com Cao [CAO17], o desenvolvimento da maturidade na extração de conhecimento em dados envolve um ciclo evolutivo de complexidade crescente que abrange aspectos de tecnologia e do negócio. Esta colocação confirma-se no estudo já citado do MITSloan [RAN18] que mostra que uma minoria das empresas consegue superar o desafio de obter conhecimento de valor a partir dos seus dados. Em valores numéricos, apenas 3% das organizações pesquisadas no estudo conseguiram vencer o citado desafio entre 2016 e 2017.

O quadro evolutivo traçado por Cao [CAO17] identifica dois grandes estados definidos a partir do nível de maturidade no aproveitamento de dados para a geração de conhecimento. O primeiro estado caracteriza organizações que usam dados como Sistemas de Apoio à Decisão (SAD). Neste caso, o autor coloca que a análise do conhecimento visa validar hipóteses predefinidas, que surgem do próprio contexto onde o conhecimento é aplicado. No segundo estado, os dados são tratados de forma a criar visões e novas hipóteses que precisam ser avaliadas, aprendidas e integradas ao conhecimento já existente no domínio ao qual pertencem. Este segundo grupo caracteriza organizações que usam modelos de KDD. O autor acrescenta que na passagem de um estado a outro existe um aumento de complexidade na forma que se realiza a análise de dados. Esta complexidade é gerada pela multidisciplinaridade e interdisciplinaridade dada por elementos de tecnologia e do negócio que devem estar corretamente elencados para realizar a transição entre os dois estados.

Já em 2011 Peterson [PET11] menciona que muitas organizações trabalham com SAD. Dentre estes sistemas ele destaca os Sistemas de *Customer Relationship Management* - CRM e *Business Intelligence* – BI. Ainda ressalta que outras abordagens como KDD utilizam-se com menor frequência devido às dificuldades que apresentam os projetos na sua implantação e à incerteza dos seus resultados.

A partir do contexto descrito, pode-se afirmar que organizações com Cultura Orientada a Dados apresentam um comportamento diferenciado, vindo da geração e aplicação do conhecimento criado a partir dos seus dados. Existem diferentes procedimentos para extrair este conhecimento. Estes procedimentos se caracterizam por uma crescente complexidade na medida que se procura extrair conhecimento de maior valor para a organização.

Este trabalho tem como objetivo definir um Método, entendendo “método” como uma sequência organizada de atividades que visa atingir um objetivo determinado. O referido Método tem como objetivo facilitar que as organizações desenvolvam um comportamento diferenciado que as caracterize como Organizações Inteligentes. Este comportamento diferenciado é propiciado pelo uso que estas organizações fazem dos dados para geração de novo conhecimento mediante a aplicação de modelos de KDD. Como requisito para aplicação do método as organizações devem usar modelos SAD, em particular soluções de BI, em pelo menos alguma das suas áreas de negócio. Na aplicação do Método estas organizações poderão fazer a transição para a aplicação de KDD atendendo a uma sequência organizada de atividades e sem que isto afete a operação da solução BI já instalada. Desta forma as funcionalidades existentes poderão ser mantidas sem gerar impactos na sua operação, ao mesmo tempo um novo potencial na forma como os dados são explorados poderá ser desenvolvido.

O ponto de partida e de chegada para a execução do Método é o Negócio. O entendimento da necessidade do Negócio em relação à obtenção do novo conhecimento direciona todas as atividades que compõem o Método. Finalizando, o Negócio avalia e valida os resultados, padrões de dados, obtidos. O Método apenas deve cobrir a entrega dos novos padrões de dados ao Negócio, não entrando em detalhes de como o negócio fará a validação dos padrões, sua aplicação e consequente acompanhamento de resultados.

O escopo para aplicação do Método envolve tanto aspectos computacionais como do negócio. Devido à amplitude do escopo, para este trabalho serão abordados apenas aspectos computacionais, vinculados ao processamento de dados: identificação de requisitos, seleção e tratamento de dados, construção do Modelo de Dados, seleção e aplicação de ferramentas para a Construção do Modelo de Dados, entrega de resultados. Reserva-se a menção de aspectos do negócio apenas como forma de contextualizar os aspectos computacionais,

também no papel de geradores de requisitos, avaliadores dos produtos gerados e fornecedores de conhecimento de apoio a todo o processo computacional.

1.1. MOTIVAÇÃO

Kimball [KIM13] define BI como um termo genérico usado para descrever um ativo organizacional empregado para reunir informação interna e externa à organização. Esta informação é usada para dar apoio à tomada de decisão. O autor acrescenta que a aplicação BI é o principal componente de um modelo de SAD que usa este tipo de abordagem de dados. Esta aplicação pode contemplar um leque bem amplo de tecnologias que abrangem das mais simples, que permitem apenas eventuais pesquisas diretas nos dados; até as mais sofisticadas, que propiciam a implementação de modelos mais complexos para a combinação e tratamento dos dados. O fato é que em qualquer uma das suas versões o objetivo principal destas aplicações é disponibilizar informações para a tomada de decisão, possibilitando aos usuários melhor e mais rápidas decisões.

Conforme Angeloni [ANG03], o processo de tomada de decisão parte de um conhecimento preexistente dos usuários que aplicam um processo seletivo sobre a informação recebida com o intuito de apoiar o processo de tomada de decisão. Assim sendo, pode-se afirmar que a análise de dados é realizada a partir de hipóteses definidas pelos próprios usuários que usam estes dados como base de testes para as referidas hipóteses.

De acordo com Fayyad [FAY96], KDD é um processo de descoberta de conhecimento útil de caráter preditivo, a partir de dados. A Mineração de Dados se refere a uma etapa determinada do processo onde aplicam-se algoritmos específicos de inteligência artificial para extrair padrões dos dados. Estes padrões de dados serão posteriormente validados para então incorporá-los ao conhecimento organizacional. Fayyad [FAY96] acrescenta que a aplicação de métodos de Mineração de Dados pode ser uma atividade nem sempre bem-sucedida, levando facilmente à descoberta de padrões inválidos.

Russom [RUS11] apresenta uma pesquisa realizada em organizações que já trabalham com soluções de BI e desejam aplicar métodos de KDD. O estudo aponta as dificuldades enfrentadas nesta transição, identificando tanto questões técnicas quanto do negócio. O autor refere-se ao KDD como “formas avançadas de análise”,

acrescentando que estas usam uma coleção de técnicas e ferramentas que devem estar perfeitamente elencadas para obter os resultados desejados. Já no negócio, o autor enfatiza o fato que o novo usuário é um analista que trabalha sobre grandes volumes de dados que precisa descobrir novo conhecimento que seja relevante para o negócio. As novas habilidades necessárias para este perfil é uma das principais barreiras para a transformação. A falta de patrocínio é considerada como uma segunda barreira. Esta pesquisa ainda apresenta alguns depoimentos muito interessantes, como por exemplo, o de um gestor da área de BI. Este gestor compara a transição para adoção de um modelo de KDD como uma “tormenta de fatores” que precisam ser elencados para tornar a aplicação de modelos de KDD possível. Em forma mais concreta, quando os participantes da pesquisa foram questionados sobre se consideravam a transição uma oportunidade ou um problema, 30% dos entrevistados afirmaram categoricamente ser um problema dado o alto grau de dificuldade implícita nesta transformação.

A constatação dos benefícios gerados pelo uso de KDD nas organizações relatados na pesquisa realizada pelo MITSloan [RAN18] é confirmada por diversos autores que apresentam estudos de caso sobre os mesmos desafios como Angera et al. [ANG16], Haas [HAA14], Gagnon [GAG15], Casali [CAS12], Hammad et al. [HAM14], Chan et al. [CHA98], entre outros.

Constitui assim a motivação para este trabalho colaborar com as organizações para que estas se tornem mais maduras no uso dos dados e possam apresentar um comportamento diferenciado que as caracterize como organizações inteligentes. Desta forma, estas organizações poderão aplicar o conhecimento proveniente da sua base de dados para cumprir sua missão.

1.2. DESENHO DA PESQUISA

Esta Seção descreve como foi estruturada a pesquisa para o desenvolvimento do presente trabalho.

A fase inicial teve caráter exploratório e documental buscando conceituar o termo “inteligência” e a forma como ela se manifesta. O estudo teve como ponto de partida a inteligência individual em suas diferentes abordagens, passando pela inteligência coletiva, para posteriormente consolidar o critério de Inteligência Organizacional, focando nas suas manifestações.

Em um segundo momento uma revisão da literatura permitiu identificar o Modelo de BI usado para definir o Estado Inicial, ponto de partida para a aplicação do Método apresentado em este trabalho.

A continuação, outra revisão da literatura foi realizada sobre modelos de KDD. Como resultado desta revisão foi possível identificar o Modelo de KDD que suportaria o Estado Final, ponto de chegada na da aplicação do Método.

No seguinte passo trabalhou-se na construção do Método e planejou-se sua avaliação. Como forma de avaliação foi elaborado um instrumento, um Tutorial. Para avaliar o Tutorial foram aplicados os questionários de avaliação de Utilidade e Facilidade de Uso que fazem parte da Modelo TAM.

A avaliação do Tutorial foi realizada em duas etapas. Na primeira etapa foi realizada uma avaliação preliminar do Tutorial e do questionário elaborado a partir do Modelo TAM. Esta avaliação foi realizada por dois especialistas. A partir desta avaliação foram realizados os ajustes pertinentes no questionário e uma revisão no Tutorial. Posteriormente o Tutorial foi submetido a avaliação de um Grupo Focal que respondeu ao questionário sobre Utilidade e Facilidade de Uso já mencionado.

Complementando a avaliação, a autora deste trabalho realizou um exercício prático de aplicação do Método. Neste exercício o Método foi aplicado sobre um conjunto de dados abertos.

1.3. ORGANIZAÇÃO DO TRABALHO

O restante deste trabalho está organizado de acordo com a descrição a seguir. O Capítulo 2 apresenta a fundamentação teórica. O Capítulo 3 apresenta o Método proposto. O Capítulo 4 apresenta os mecanismos usados para avaliação do Método. O Capítulo apresenta um exercício de aplicação do Método. O Capítulo 6 apresenta os trabalhos relacionados. Finalmente, o Capítulo 7 apresenta as considerações finais e os trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Este Capítulo apresenta a base teórica referente aos principais conceitos que envolvem este trabalho. A Seção 2.1 apresenta um resumo introdutório do conceito de Inteligência. A Seção 2.2 apresenta o conceito de Inteligência Individual. A Seção 2.3 apresenta o conceito de Inteligência Coletiva e Organizacional. A Seção 3.4 apresenta os fundamentos do processo que transforma informação em conhecimento. A Seção 2.5 apresenta o Modelo de Kimball. A Seção 2.6 apresenta o processo de KDD. Finalmente a Seção 2.7 apresenta Modelos de KDD.

2.1. O CONCEITO DE INTELIGÊNCIA

O conceito “inteligência” categoriza pessoas e grupos como um fator diferencial de alto valor. O termo tem sido foco de estudo desde a antiga Grécia, abordado e pesquisado por diferentes áreas de conhecimento. Diante da vastidão dos trabalhos publicados por inúmeros autores em diferentes épocas, pode-se conjecturar que dificilmente existirá uma teoria consensual sobre o que é inteligência. As inúmeras conexões entre as diferentes teorias nas mais variadas áreas de conhecimento e sua evolução descontínua na linha do tempo respaldam o prognóstico. Consequentemente, caracterizar a forma como se identifica, entre indivíduos ou grupos, um comportamento inteligente torna-se uma tarefa árdua.

Já na abordagem aplicada no coletivo, mais especificamente no contexto das organizações, o conceito “organizações inteligentes” também é abordado e discutido por um grande número de autores. Na definição do termo, diversos autores elencam diferentes elementos organizacionais para descrever organizações que se enquadram dentro desta classificação. Estes elementos comportam algumas características comportamentais das organizações vinculadas à obtenção de seus objetivos.

2.2. INTELIGÊNCIA INDIVIDUAL

O que faz uma pessoa ser reconhecida como inteligente? Pode ser sua competência em responder em forma rápida e assertiva a diferentes estímulos ou de transitar e agir com propriedade em ambientes de alta complexidade, ou ainda a sua

competência para abstração e reuso de conhecimento em diferentes contextos, assim como a alta expertise em alguma área de conhecimento em particular. Seja qual for seu comportamento, estas pessoas são identificadas, diferenciadas e valorizadas dentro do seu contexto físico e temporal de atuação. A inteligência é reconhecida como uma qualidade humana desejável, notória, universal e atemporal.

Se popularmente o termo “inteligência” é amplamente compreendido, desde o rigor da ciência a inteligência é um dos temas mais polêmicos e controversos.

De acordo com Gardner [GAR98], a inteligência pode ser estudada em muitos níveis que vão desde o sistema neuronal até o comportamento social. O estudo da inteligência não pode ser realizado dentro de uma única disciplina. Portanto, qualquer tentativa de compreender a inteligência precisa incluir várias disciplinas e diferentes autores dentro delas.

Virgolim [VIR14] faz uma ampla análise do termo sob a ótica da psicologia e da filosofia. A autora inicia sua análise colocando que o desenvolvimento do conceito de “aquilo” que chamamos de inteligência é bastante antigo. Virgolim cita diferentes autores desde Platão (428 a.c. – 347 a.c.) até teorias contemporâneas, passando por Descartes (1596 - 1650), Locke (1632 - 1704), Kant (1724 - 1804), Piaget (1896 - 1980) entre muitos outros. Finalmente, a autora conclui que é difícil argumentar que existe uma definição que pode ser considerada a correta para o termo “inteligência”. No entanto, algumas definições são claramente mais concisas, outras menos precisas ou gerais. Mas, existe um consenso que “inteligência ou comportamento inteligente” possui uma série de atributos:

- É uma propriedade que caracteriza o comportamento de um indivíduo dentro do seu contexto de atuação.
- Depende da capacidade de adaptação e respostas do indivíduo em novos contextos.
- Está relacionada com a capacidade de o indivíduo ter sucesso em relação a um objetivo.
- Está totalmente vinculada à capacidade de aprender do indivíduo.

2.3. INTELIGÊNCIA COLETIVA E INTELIGÊNCIA ORGANIZACIONAL

A inteligência é normalmente percebida como uma capacidade individual, porém se refletir-se em problemas complexos como a construção de um edifício, pensar na inteligência em caráter individual torna-se complicado, evidenciando-se assim seu caráter coletivo. Na construção de um software deve-se pensar no arquiteto que procurará a melhor combinação de diferentes Hardware e Software para atender à solicitação, o desenvolvedor que escreverá o código, o testador que validará o produto, o analista de negócios identificará os requisitos, etc. Todos eles reúnem um conjunto de expertises necessárias ao objetivo final. A reunião certa de todas estas expertises garante a correta construção do software. Em uma forma mais ampla, pode-se afirmar que quanto melhor os grupos humanos conseguem se construir em coletivos inteligentes, maior a sua chance de sucesso.

Segundo Pierre Levy [LEV00] a inteligência coletiva se dá na união de saberes, experiências e ideias. Deve ser compreendida com a expressão “trabalhar em comum acordo”.

Schvarstein [SCH03] cita Williamsen e Colledge (1995) para definir a inteligência social como “habilidades e capacidades necessárias para criar e manter a comunidade”. Entendendo comunidade como “um grupo específico de pessoas que costumam viver em uma área geográfica determinada, que compartilhem uma cultura comum, que estão organizadas em torno de uma estrutura social e que mostram uma certa consciência da sua identidade como grupo”. Já no contexto organizacional, Schvarstein citando Morgan(1986) pontua que a inteligência seja uma qualidade individual, uma organização também pode se autodeclarar inteligente se for capaz de promover reflexões dos seus integrantes sobre a adequação das suas ações de acordo com os requisitos que as diferentes situações exigem.

Wilensky[WIL15] reúne os principais elementos e descobertas de pesquisas sobre organizações em meados de 1960 e conclui que: o conhecimento determina parcialmente as decisões organizacionais e seus resultados. Ele chama esse conhecimento de "inteligência", e acrescenta que a inteligência organizacional se traduz na capacidade de processar, interpretar e comunicar a informação. Em outras palavras, a inteligência organizacional está relacionada aos recursos que uma organização dedica, os tipos de especialistas que ela utiliza, as funções que esses

especialistas desempenham e a relação da organização com seu ambiente externo e interno.

McMaster [MCM16] define a “Organização Inteligente” como a organização que tem a capacidade de reunir, gerar e disponibilizar informação para construir o conhecimento que lhe fornece subsídios para um determinado comportamento.

CHOO [CHO06] reforça o vínculo entre informação e comportamento inteligente das organizações afirmando que a informação é um componente intrínseco de quase tudo o que a organização faz. A clara percepção dos processos que transformam informação em conhecimento e posteriormente em ação faz com que as organizações sejam capazes de perceber a importância de suas fontes tecnológicas de informação.

Muller [MUL12], na sua análise, cita Albrecht (2003). Este coloca que a Inteligência Organizacional pode ser compreendida como “a capacidade que uma organização possui de mobilizar todo o seu poder cerebral e concentrar tal poder na conquista de sua missão”. Já citando Cruz e Dominguez (2007), estes definem a Inteligência Organizacional como uma capacidade desenvolvida pelo aprendizado organizacional sistêmico, que permite à empresa perceber adequadamente os ambientes interno e externo por meio do processamento das informações obtidas nestes ambientes, além de gerar novos conhecimentos.

Perante o exposto, conclui-se que organizações inteligentes apresentam um comportamento diferenciado gerado pela aplicação de conhecimento, produto do processo de transformação da informação.

2.4. DA INFORMAÇÃO AO CONHECIMENTO

A informação cumpre um papel fundamental no desenvolvimento das organizações, ela se transforma em um verdadeiro vetor que direciona as ações empresariais. A forma como se executa a gestão da informação depende da tecnologia disponível. O Instituto MITSloan [KIR13], já referenciado anteriormente, define Organizações com Cultura Orientada a Dados como organizações com padrões de comportamento e práticas nas quais se compartilha a crença de que a criação, entendimento e aplicação de determinado tipo de dados e informação possui um papel crítico no sucesso da organização.

A criação de conhecimento organizacional não é um assunto novo. Este tópico é abordado principalmente pela área de Gestão do Conhecimento e conta com reconhecidíssimos autores como Nonaka, Takeuchi, Davenport, Senge e Garvin, entre outros. Paulo Yazig Sabbag [SAB07], em seu livro “Espirais do Conhecimento”, faz uma análise de diferentes teorias apresentadas por alguns autores consagrados que descrevem como o conhecimento é gerado dentro das organizações. Sabbag coloca que a formação do saber (conhecimento) não é um processo instantâneo. Como processo, está constituído por um número determinado e consecutivo de passos que levam desde o saber individual, próprio de cada indivíduo, à generalização, propriedade de grupos. A visão de processo também determina sua dimensão temporal e não instantânea. Este processo é repetitivo e ininterrupto onde a última etapa se elenca com a primeira da sua nova execução, lembrando assim de uma espiral. Diferentes autores analisados pelo autor conferem nomes variados à cada uma das atividades que integram o processo de geração de conhecimento. Porém, todos eles são unânimes ao identificar uma atividade ou um conjunto de atividades nas quais novas informações ou novos conhecimentos são absorvidos pelo indivíduo e posteriormente incorporados ao conhecimento já existente para compor um novo conhecimento. Davenport e Pruzak (1998) chamam esta atividade de Criar Conhecimento; Salim (2002) a decompõe em duas atividades chamadas de Mapear Conhecimento e de Criar Conhecimento; Sidanmaanlakka(2002) também considera duas atividades, referindo-se a elas como, Capturar Conhecimento e Criar Conhecimento.

Nonaka e Takeuchi [NON08] são autores muito referenciados quando o tema é formação de conhecimento organizacional. Estes autores descrevem o ciclo de formação de conhecimento também usando a analogia do espiral. Este ciclo é chamado de SECI, sigla que identifica cada uma das atividades que compõem o processo de formação de conhecimento: Socialização, Externalização, Combinação e Internalização. O processo de Combinação trata da união de novos conhecimentos para que durante a Internalização estes se incorporam ao conhecimento já existente no indivíduo, formando um novo.

Conhecimento é um recurso organizacional que sustenta o comportamento inteligente da organização. A criação de conhecimento é contínua, ininterrupta e elaborada pelos indivíduos.

Conforme já exposto Fayyad [FAY96] refere-se a KDD como um processo de descoberta de conhecimento útil de caráter preditivo, a partir de dados. Este processo gera padrões de dados que posteriormente serão validados pelo negócio para incorporá-los ao conhecimento organizacional. Assim sendo, as práticas de KDD possibilitam incorporar um volume maior de conhecimento ao processo de formação de conhecimento organizacional, tornando-o mais rico e dinâmico. Desta forma cresce a “força mental” da organização para que esta possa manifestar seu comportamento de forma “inteligente”.

2.5. BI E SEUS MODELOS

De acordo com Breslin [BRE04], um *Data Warehouse* contém grandes quantidades de dados detalhados de séries temporais usados para dar suporte à tomada de decisão. As fontes da maioria dos dados em um *Data Warehouse* são sistemas internos de processamento de transações (também conhecidos como sistemas operacionais). Um software especializado resume, reconcilia e manipula os dados originais antes destes serem disponibilizados para armazenamento e uso no *Data Warehouse*. Uma grande variedade de ferramentas está disponível no mercado para acesso, design e manutenção destes dados, formando assim um modelo SAD. A autora afirma que apesar da variedade de escolhas de ferramentas e metodologias, existem apenas dois modelos que dão a fundamentação básica para a composição do SAD: Inmon e Kimball.

Breslin [BRE04] ainda faz um breve relato histórico justificando o motivo pelo qual estes autores consideram-se os “gigantes” dentro da sua área de atuação. Sem dúvida Ralph Kimball e William H. Inmon, são os autores mais consagrados quando o tema é modelos de SAD. Inúmeros artigos já foram publicados onde seus trabalhos são referenciados, comparados e destacados como marcos dentro da evolução e desenvolvimento da disciplina. A Autora realizou uma análise comparativa entre os modelos, ela explica que existem poucas semelhanças entre as duas abordagens e enfatiza que as diferenças entre os dois modelos são muitas e profundas. A Tabela 1 e a Tabela 2 apresentam quadros comparativos entre os modelos estudados. A Tabela 1 apresenta um estudo comparativo sobre os seguintes aspectos dos modelos: metodologia, arquitetura, modelagem de dados e tipo de abordagem. Já a Tabela 2 faz um comparativo dos aspectos positivos de ambos modelos quando estes abordam

os seguintes pontos: natureza dos requisitos para suporte à tomada de decisão, requisitos de integração de dados, estrutura de dados, escalabilidade, persistência de dados, requisitos de pessoas, tempo de entrega e custo de implantação.

Tabela 1: Tabela comparativa dos modelos de Inmon e Kimball com foco na metodologia e Arquitetura, BRESLIN [BRE04], traduzida pela Autora

	Kimball	Inmon
Metodologia e arquitetura		
Abordagem geral	Bottom-up	Top-down
Estrutura de arquitetura	<i>Datamarts</i> modelam um único processo de negócio; a consistência da empresa é atingida através de bus de dados e dimensões conformadas.	Data Warehouse (atômica) de toda a empresa "alimenta" as bases de dados dos departamentos
Complexidade do método	Bem simples	Bastante complexo
Comparação com metodologias de desenvolvimento estabelecidas	Processo de quatro etapas; parte dos métodos RDBMS	Derivado da metodologia espiral
Discussão do desenho físico	Leve	Bastante completo
Modelagem de dados		
Orientação dos dados	Por processo	Por assunto ou data-driven
Ferramentas	Modelagem dimensional; parte da modelagem relacional	Tradicionais (ERDs, DISs)
Acessibilidade para o usuário final	Alta	Baixa
Filosofia		
Audiência primária	Usuário final	Profissionais de TI
Lugar na organização	Transformador e retentor de dados operacionais	Parte integral da CIF (Fábrica de Informação Corporativa)
Objetivo	Entregar uma solução que permita que os usuários finais possam solicitar diretamente os dados e ainda obter tempos de resposta razoáveis	Entregar uma solução técnica baseada em métodos de base de dados e tecnologias comprovados

Tabela 2: Tabela comparativa dos modelos de Inmon e Kimball com foco na natureza do processo de tomada de decisão, BRESLIN [BRE04], traduzida pela Autora

Característica	Favorece Kimball	Favorece Inmon
Natureza dos requisitos de suporte de decisão da organização	Tática	Estratégica
Requisitos de integração de dados	Áreas de negócio individuais	Integração de toda a empresa
Estrutura de dados	Métricas de negócio, mensuração de performance e <i>scorecards</i>	Dados não métricos e para dados que serão aplicados para múltiplas e variadas necessidades de informação
Escalabilidade	Precisa adaptar-se para necessidades voláteis em um escopo limitado	Escopo crescente e requisitos de mudança são críticos
Persistência de dados	<i>Source systems</i> são relativamente estáveis	Alta taxa de mudança dos <i>Source systems</i>
Requisitos de pessoas e habilidades	Times pequenos de generalistas	Maiores times de especialistas
Tempo de entrega	Necessidade do primeiro data warehouse é urgente	Os requisitos da organização permitem um tempo de start-up mais longo
Custo de implantação	Custos de start-up mais baixos, com cada projeto subsequente custando aproximadamente o mesmo	Custos de start-up mais altos, com projetos subsequentes mais baratos

Breslin ressalta que a simplicidade do modelo de Kimball permite um maior grau de integração entre a TI e o negócio, garantindo a participação do usuário final. Ainda acrescenta que Kimball usa como base para sua abordagem necessidades específicas do negócio. Esta abordagem se concretiza na modelagem usada para os dados que é totalmente orientada aos processos do negócio

Em particular, este trabalho considera o modelo de Kimball. Esta escolha se deve à abrangência e profundidade com que o autor trata do tema. A escolha se vê reforçada pela análise comparativa entre os modelos realizada por Breslin [BRE04] que destaca características do modelo escolhido como: simplicidade do desenho, orientação ao usuário final.

2.5.1. O Modelo de Kimball

Kimball [KIM13] define BI como um termo genérico, empregado para descrever um ativo organizacional usado para reunir informação interna e externa à organização, aplicada para dar apoio à tomada de decisão. Acrescenta ainda, que o termo BI surge em 1990, referindo-se à análise de dados armazenados no *Data Warehouse* - DW.

Outros autores referem-se apenas ao BI como a ferramenta de suporte ao processamento da informação. Kimball chama este binômio de DW/BI. Para este binômio, o autor define um modelo, que sob o rótulo de “ciclo de vida”, pode ser usado tanto para a implantação da nova tecnologia quanto para qualquer tipo de manutenção de seus componentes após feita a implantação. Este modelo é formado por diferentes elementos que, na execução do ciclo de vida, elencam atividades que passam a integrar um projeto. Os elementos que integram o modelo de Kimball estão ilustrados na Figura 1 e descritos a seguir:

- Planejamento do Projeto: para cada iteração de aplicação do modelo, ciclo de vida, é definido um projeto como forma de planejar e gerenciar todo o processo de implantação e/ou aplicação de mudanças que precisam ser implementadas nos elementos constitutivos do modelo.
- Gerenciamento do Projeto: uma vez definido o projeto a sua gestão permite acompanhar e execução garantido que esta termine com sucesso.
- Definição dos Requisitos do Negócio: o êxito do modelo está fortemente ligado ao entendimento dos requisitos do negócio. Este entendimento garante que o fluxo de informação seja constante e adequado às necessidades dos usuários.
- Projeto da Arquitetura Técnica: a plataforma do DW/BI requer numerosas tecnologias. Para determinar a tecnologia necessária três fatores devem ser considerados: requisitos do negócio, atual contexto técnico e diretrizes estratégicas.
- Seleção e Instalação de Produtos: uma vez determinada a arquitetura, os produtos para compor esta arquitetura devem ser selecionados, instalados e testados.
- Modelagem Dimensional: quando identificam-se os requisitos do negócio, também identificam-se as necessidades de informação de acordo com os processos chave do negócio. Desta forma pode-se garantir que o fluxo de dados chegue a todos os pontos do negócio onde são necessários.
- Projeto Físico: o Projeto Físico ou *database* foca nas estruturas físicas, incluindo o banco de dados e o contexto que vai garantir sua segurança. Estratégias de desempenho, particionamento, agregação e *On-line Analytical Processing* - OLAP identificam-se neste momento.

- Projeto e Desenvolvimento de ETL(*Extract Transform Load*): o ETL implementa o desenho e desenvolvimento da extração, transformação e carga de dados.
- Especificação da Aplicação Analítica: identifica as ferramentas de BI que atendem aos requisitos do negócio e as melhores interfaces que atendem às necessidades dos usuários.
- Desenvolvimento da Aplicação Analítica: Seguidamente à escolha da ferramenta e de acordo com a mesma os metadados precisam ser configurados assim como a infraestrutura para receber a ferramenta.
- Implantação: garante que todos os elementos do modelo sejam corretamente elencados e conjuntamente validados para garantir a correta operação da solução.
- Manutenção e Expansão: a plataforma DW/BI deve estar pronta para poder expandir-se ou mudar para melhor atender às necessidades do Negócio.

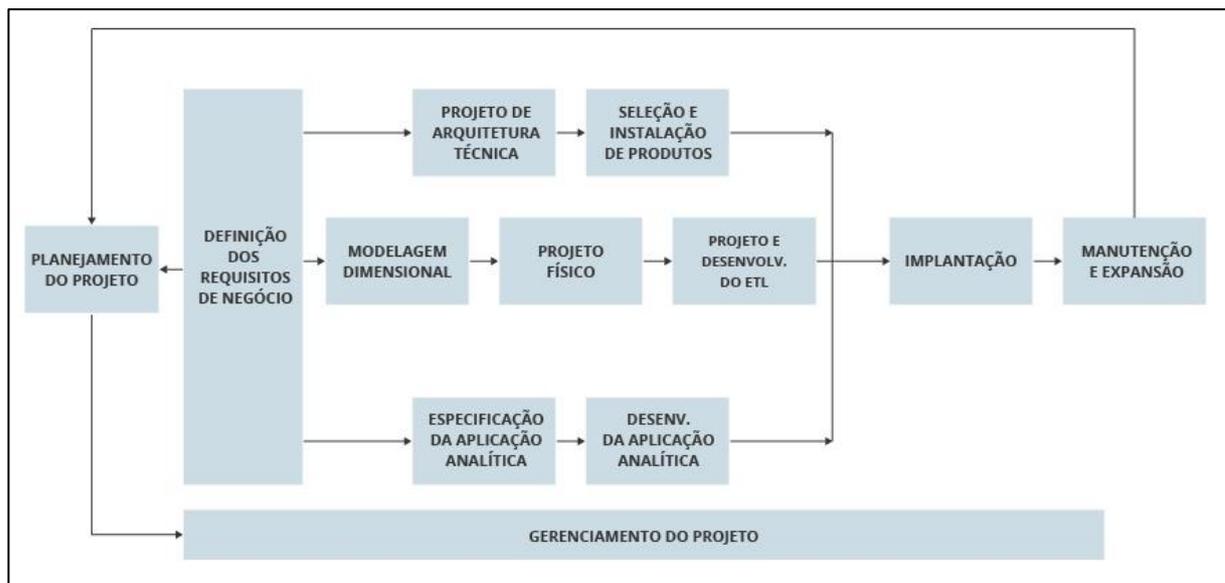


Figura 1 - Modelo de BI, Kimball [KIM13], traduzido pela autora.

2.6. O PROCESSO DE KDD

De acordo com os autores do Modelo CRISP-DM[CHP00] o processo de extração de conhecimento a partir de grandes conjuntos de dados é um processo não trivial, interativo e iterativo. O processo de KDD é rotulado como interativo, indicando a necessidade da atuação eventual de um especialista como responsável do controle

do processo. O especialista aplica os recursos computacionais e do Negócio em função dos fatos observados e dos resultados envolvidos. O termo iterativo indica a possibilidade de repetições das atividades do processo em busca de resultados satisfatórios.

Os autores acrescentam que a aplicação do processo de KDD permite identificar Padrões. Estes devem ser interpretados por humanos para se converterem em conhecimento. Um Padrão descreve fatos ou tendências associadas a um conjunto de dados. Os Padrões podem ser classificados em 2 tipos básicos: preditivos e descritivos. Padrões preditivos são construídos com o intuito de resolver o problema específico de prever valores de um ou mais atributos. Os Padrões descritivos apresentam informações que um especialista do domínio da aplicação pode não conhecer. De acordo com os autores, os Padrões encontrados como o resultado da aplicação do processo de KDD devem atender a 4 critérios. Eles devem ser:

- Compreensíveis: permitem um entendimento claro pelos usuários do domínio de aplicação, isto implica uma representação simples que os torne inteligíveis. Uma técnica possível para atingir este objetivo é a visualização.
- Válidos: o conhecimento deve ser adequado ao contexto de aplicação.
- Novos: devem acrescentar novos conhecimentos aos já existentes no contexto de aplicação e principalmente para o especialista do domínio.
- Úteis: podem ser aplicados de forma a proporcionar benefícios ao contexto da aplicação do KDD.

2.7. MODELOS DE KDD

De acordo com pesquisa realizada por Mariscal [MAR10], existem apenas dois modelos seminais para KDD, o Modelo de Fayyad e o Modelo CRISP-DM. Os outros modelos existentes são apenas híbridos dos anteriores ou aplicações práticas dos mesmos, conforme ilustrado na Figura 2.

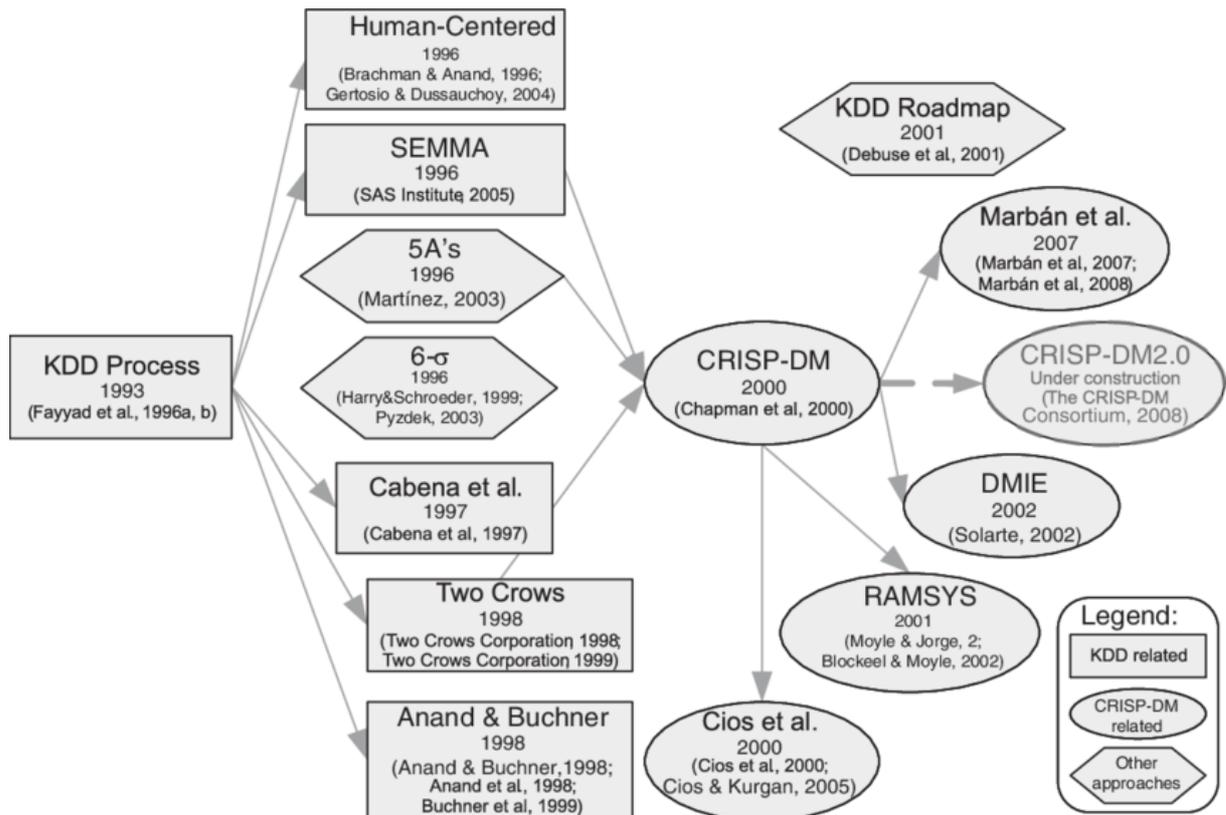


Figura 2 - Modelos de KDD de acordo com a pesquisa de Mariscal [MAR10]

Trabalhos publicados posteriormente ao trabalho de Mariscal demonstram que os modelos de Fayyad e CRISP-DM continuam sendo seminiais. Estes trabalhos usam os referidos modelos como base para sugestão de melhorias, enfatizando em alguns aspectos onde são considerados deficientes, ou tratam de aplicações práticas onde considerou-se necessário fazer algum tipo de adaptação para garantir a aplicabilidade dos mesmos. Vale destacar alguns destes trabalhos.

Fasseler [FAS17] apresenta o caso do modelo CRISP-DM, descrevendo detalhadamente a aplicação de cada uma das suas etapas. O autor destaca a flexibilidade do modelo que permite intercalar as fases de execução de acordo com os resultados obtidos após a aplicação de cada uma delas.

O trabalho de Rahman [RAH14] foca na qualidade de dados associada à preparação dos mesmos dentro do contexto do modelo CRISP-DM. O autor avalia o impacto do processo de limpeza de dados associado à qualidade dos resultados da aplicação do modelo.

Hammad et al. [HMM14] ressalta que é comum em projetos de construção industrial que os dados sejam coletados e descartados sem serem analisados para extrair conhecimento útil. A partir de um caso prático, o autor mostra a aplicabilidade

do método, sobre um *data warehouse*, como forma de aproveitar o conhecimento existente nos dados. O modelo foi aplicado a três estudos de caso diferentes, usados para validar a aplicabilidade do mesmo.

Niaksu [NIA15] desenvolve um novo modelo, originário do CRISP-DM, considerando sua aplicação prática na Medicina. Este novo modelo é chamado de CRISP-MED-DM. As onze etapas do novo modelo proposto apresentam uma abordagem iterativa para definir o conjunto de dados ideal e encontrar o algoritmo de *Data mining* com melhor desempenho na sua aplicação.

Outro trabalho anterior à revisão Mariscal foi apresentado por Sharma em 2008 [SHA08]. O autor usa como base o modelo CRISP-DM para definir um novo Modelo chamado IKDDM (*Integrated Knowledge Discovery Data Model*). Sharma ressalta a falta de uma descrição mais detalhada das atividades do CRISP-DM e como elas se vinculam na interação entre as diferentes fases do modelo. Com esta abordagem ele apresenta esta nova versão que pretende cobrir esta deficiência.

2.7.1. O Modelo de Fayyad

Fayyad et al. [FAY96] descreve KDD como um processo. Este processo envolve um conjunto de etapas que incluem desde a forma como os dados são armazenados até a interpretação dos resultados obtidos. Este processo é altamente iterativo, envolvendo assim o usuário no andamento das diferentes etapas. A Figura 3 apresenta o processo definido por Fayyad et al. seguido da breve definição de cada uma das etapas que o constituem

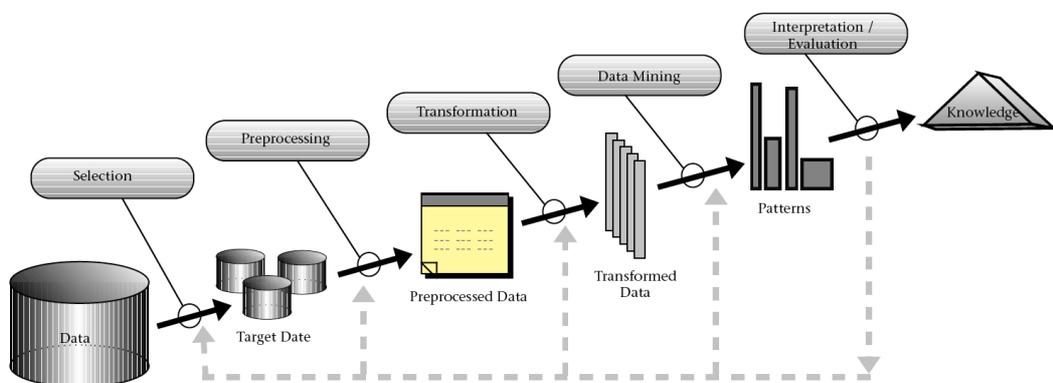


Figura 3 - Processo de KDD, Fayyad et al. [FAY96]

- Etapa I - *Selection*. Consiste em desenvolver uma compreensão do domínio de aplicação e identificar o objetivo do processo KDD do ponto de vista do cliente. Isto permite selecionar um conjunto de dados ou focar um subconjunto de variáveis ou amostras de dados sobre o qual a descoberta deve ser realizada.
- Etapa II – *Preprocessing*. Consiste na limpeza e preparação dos dados. As operações básicas incluem a remoção de ruído e estratégias para lidar com os campos de dados em falta. Também é realizada a redução de dados que consiste em eliminar dados invariantes ou variáveis não significativas.
- Etapa III – *Transformation*. Após serem selecionados, limpos e pré-processados os dados necessitam ser armazenados e formatados adequadamente para que os algoritmos de aprendizagem de máquina possam ser aplicados.
- Etapa IV - *Data Mining*. Consiste em associar aos objetivos do processo KDD (Etapa I) um método particular de mineração de dados que será usado para a pesquisa dos padrões. Por exemplo: a classificação, a regressão, o agrupamento, etc.
- Etapa V – *Interpretation/Evaluation*. Consiste na interpretação dos padrões extraídos, podendo retornar à alguma das etapas anteriores para uma iteração adicional. Esta etapa também pode envolver a visualização dos padrões e modelos extraídos

2.7.2. O Modelo CRISP-DM

O modelo CRISP-DM [CHP00] (*Cross Industry Standard Process for Data Mining*) implementa um ciclo composto por um conjunto de tarefas, organizadas em diferentes etapas, e os relacionamentos entre elas. As relações podem existir ou não dependendo dos objetivos, os interesses do usuário e dos dados. Este ciclo consiste em seis etapas, conforme apresentado na Figura 4.

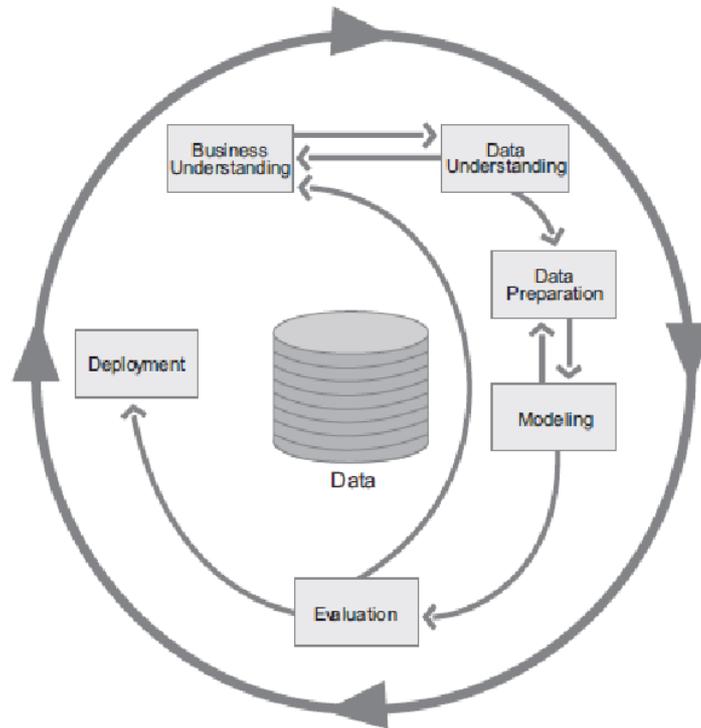


Figura 4 - Modelo CRISP-DM [CHA00]

A sequência de execução das etapas não é pré-determinada. O resultado da execução de cada etapa determina a etapa que será executada a continuação. A sequência das setas indica as dependências mais importantes e as mais frequentes. O processo não termina quando a solução é obtida. A avaliação dos resultados pode indicar a necessidade de execução de um novo ciclo. As etapas que compõem o processo são descritas brevemente a seguir.

- *Business Understanding*. Esta etapa inicial concentra-se na compreensão dos objetivos requisitos do projeto de uma perspectiva de negócios, em seguida, na conversão em um plano preliminar projetado para atingir os objetivos.
- *Data Understanding*. Esta etapa começa com a coleta de dados inicial e prossegue com atividades que permitem familiarizar-se com os dados, identificar problemas de qualidade, descobrir primeiro insights e/ou detectar subconjuntos interessantes para formular hipóteses sobre informações faltantes.
- *Data Preparation*. Esta etapa abrange todas as atividades necessárias para construir o conjunto de dados final a partir dos dados brutos. As tarefas de preparação de dados provavelmente serão realizadas várias vezes.

- *Modeling*. Nesta etapa, selecionam-se as técnicas que serão usadas para a mineração de dados. Normalmente, existem várias técnicas para o mesmo tipo de problema de mineração. Algumas técnicas têm requisitos específicos sobre a forma dos dados. Portanto, muitas vezes é necessário voltar para a fase de preparação.
- *Evaluation*. Nesta etapa é construído o modelo de dados antes de proceder à implantação final. É importante revisar minuciosamente e revisar as etapas executadas para criá-lo, em busca da certeza de que o modelo atinge adequadamente os objetivos de negócios.
- *Deployment*. Nesta etapa o novo conhecimento é apresentado ao usuário final para sua avaliação.

2.7.3. Considerações

Uma revisão da literatura realizada como parte da pesquisa que deu origem a este trabalho confirma que existem inúmeros artigos publicados onde os modelos de Fayyad e CRIP-DM são referenciados. Parte destes trabalhos propõem mudanças nestes modelos visando a solução de algum ponto considerado deficiente pelos seus autores. Outros trabalhos também propõem mudanças visando a aplicabilidade dos modelos a diferentes contextos. Estes fatos demonstram que os modelos são seminiais. Ainda, como resultado da revisão da literatura já mencionada, encontrou-se farta documentação em relação ao Modelo CRIP-DM. Esta documentação, criada por um consórcio de usuários, descreve passo a passo as ações necessárias para a aplicação prática do modelo, abordando tanto aspectos computacionais como do negócio.

Face ao exposto, o modelo CRISP-DM foi escolhido para a elaboração do Método.

3. MÉTODO PARA DESENVOLVIMENTO DA INTELIGÊNCIA ORGANIZACIONAL EM ORGANIZAÇÕES ORIENTADAS A DADOS

Conforme descrito, o Método tem como objetivo realizar a transição entre duas situações ou estados, caracterizados pela forma como o conhecimento é extraído dos dados. Porém, esta transição não deve invalidar a capacidade operacional do BI em uso e sim acrescentar novos elementos e funcionalidades para viabilizar a aplicação do modelo de KDD. Cada estado é definido por um conjunto de elementos com qualidades e características particulares. Para definir a estado inicial consideram-se os elementos que constituem o modelo de BI de Kimball [KIM13]. Para definir a situação final consideram-se os elementos que constituem o modelo de KDD, CRISP-DM [CHP00]. Pela amplitude do escopo, apenas aspectos computacionais serão considerados para a elaboração do Método.

Este Capítulo apresenta uma descrição de como foi elaborado o Método proposto e está organizada da seguinte forma: a Seção 3.1 apresenta a estrutura do Método. A Seção 3.2 apresenta como o Método foi desenvolvido. Finalmente, a Seção 3.3 apresenta a conclusões.

3.1. ESTRUTURA DO MÉTODO

O Método está constituído por um conjunto de tarefas, atividades e ações que podem vir a ser executadas sob certas premissas e/ou restrições de forma de promover ações que executem a transição entre um determinado estado inicial e outro desejado. Desta forma, estabelece-se uma correspondência entre os elementos que constituem o ponto de partida e o de chegada.

Iniciando o desenvolvimento do Método foi construída uma tabela, Tabela 3, onde se estabeleceu a correspondência entre os elementos que compõem cada um dos estados. Para cada estado identificaram-se os elementos constitutivos dos correspondentes modelos, pontos de partida e de chegada.

A partir da Tabela 3 foi possível avaliar que os elementos do estado inicial se encontram enquadrados em duas situações diferentes: não sofrem nenhum tipo de alteração ou podem ser impactados pela transição. Também se verificou que existe um conjunto de novos elementos que não fazem parte do modelo de partida, mas implementam o modelo final. Considerando estas três situações foi criado um

diagrama evolutivo do Modelo de Kimball, apresentado na Figura 5, que se traduz da seguinte maneira: os componentes em azul representam os elementos do modelo de Kimball que serão impactados pela transformação; em cor salmão, destaca-se um novo conjunto de elementos que devem ser associados aos elementos que já formam o modelo de Kimball. Finalmente, em cor cinza, estão representados os componentes que permanecem inalterados.

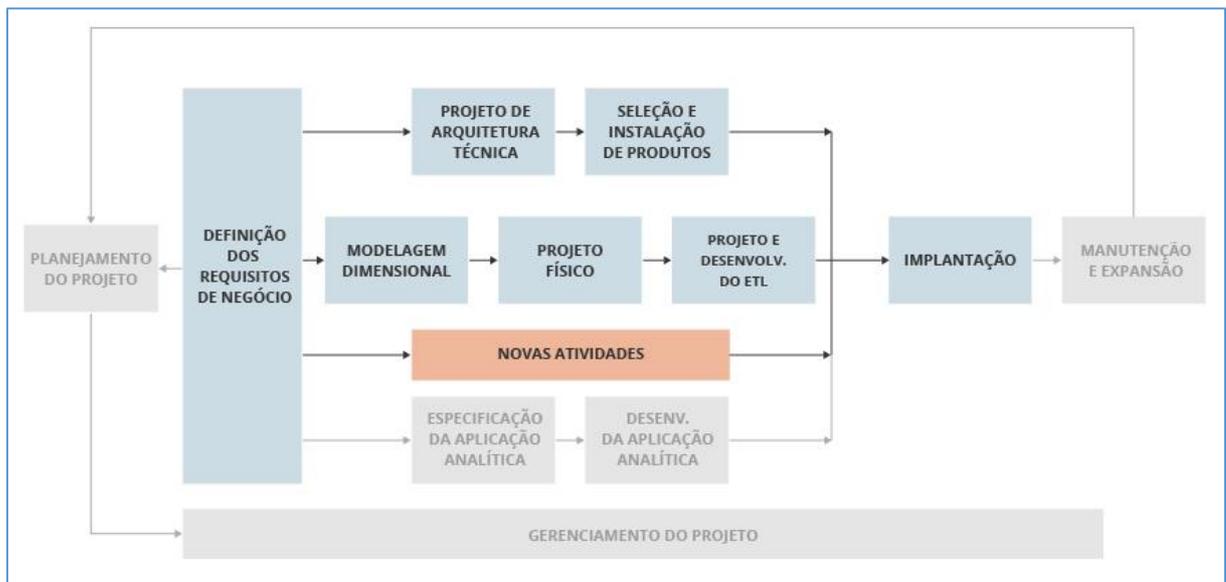


Figura 5 - Diagrama evolutivo do modelo de Kimball [KIM13], definido pela autora

Deve-se ressaltar que entre os elementos do estado final existem algumas atividades que, se bem são consideradas como novas, já estão sendo executadas em forma semelhante no estado inicial. Porém, comparando estas atividades elas são executadas tendo como base premissas diferentes e objetivos também diferentes. Assim sendo, a execução destas atividades leva à obtenção de resultados diferentes. Como exemplo pode-se mencionar a limpeza de dados. A limpeza de dados no estado inicial, acontece tanto no processo de ETL, *carga de dados no Data Warehouse*, como para a criação do Modelo de dados e tem como um dos seus objetivos eliminar os *outliers* para não gerar distorções nos resultados que serão apresentados ao negócio. Os requisitos do negócio geram premissas para o procedimento de limpeza de dados. Enquanto isso, para a limpeza de dados no estado final conservar os *outliers* pode ser interessante para obter padrões inovadores. Oferecer padrões com conhecimento novo, inovador, é uma das premissas do Negócio para o processo de KDD. Para este tipo de atividades será necessário avaliar, em tempo de execução do

Método, se serão considerados os resultados obtidos no modelo de partida ou se serão refeitos os procedimentos de acordo com o novo modelo, desconsiderando os resultados obtidos anteriormente. A decisão estará vinculada às premissas e o objetivo estabelecido para cada uma destas atividades.

A Tabela 3 apresenta a análise dos elementos dos estados inicial e final na composição do Método, indicando o tipo de transformação que os elementos do estado inicial sofrem na transição. A primeira coluna, ESTADO INICIAL, apresenta os elementos do estado inicial, componentes do modelo de Kimball. Para estes elementos estabelece-se a restrição de não se explicitar aqueles que não sofrem nenhum tipo de impacto na aplicação do Método. Isto deve-se a que estes elementos não acrescentam nenhum tipo de informação para o desenvolvimento do Método. A segunda coluna, ESTADO FINAL, apresenta os elementos que compõem o modelo CRISP-DM. Na coluna STATUS os elementos assinalados como “IMPACTADO” são os elementos do estado inicial que sofrem ou podem sofrer algum tipo de impacto pela aplicação do Método. O tipo de impacto gerado, encontra-se descrito na coluna DESCRIÇÃO. Os novos elementos contemplados pelo Método proposto encontram-se identificados com a legenda “NOVO” na coluna STATUS, e apresentam na quinta coluna, DESCRIÇÃO, uma breve descrição do seu objetivo. Os elementos com a legenda “NOVO/AVALIAR” na coluna STATUS são aqueles elementos que representam atividades que também são executadas em forma semelhante no estado inicial e precisam ser avaliadas para decidir se serão usados os resultados obtidos no estado inicial ou se serão refeitas as atividades de acordo com os novos procedimentos introduzidos pelo Método. Em certa forma, os elementos que se encontram nesta situação no estado inicial também podem ser considerados como impactados na aplicação do Modelo. A título ilustrativo a coluna SLIDE apresenta uma referência ao(s) *slide(s)* do Tutorial, onde estão detalhadas informações do Método em relação a cada um dos elementos. Cabe destacar que para a construção do usou-se como referência o Método já estruturado apresentado na Seção 3.2. Os *slides* do Tutorial podem ser encontrados no Apêndice A deste trabalho.

Tabela 3 - Análise dos elementos dos estados inicial e final na composição do Método

ESTADO INICIAL		ESTADO FINAL		STATUS	OBSERVAÇÃO	SLIDE
DEFINIÇÃO DOS REQUISITOS DO NEGÓCIO	Identificar requisitos do Negócio	Sem atividades correspondentes		IMPACTADO	Os requisitos do Negócio são identificados em forma diferente assim como os critérios de sucesso para sua avaliação	15,16
	Identificar participantes			IMPACTADO	Outros integrantes do negócio envolvidos na identificação de requisitos devem ser identificados domínios de conhecimento	
PROJETO DE ARQUITETURA TÉCNICA	Fontes de Dados			IMPACTADO	Impacto gerado pela introdução de novos dados	17,18
	Metadados			IMPACTADO	Impacto gerado pela introdução de novos aplicativos e dados	
	Sistema de ETL			IMPACTADO	Impacto gerado pela introdução de novos dados	
	Servidor de Apresentação			IMPACTADO	Impacto gerado pela introdução de novos aplicativos	
	Aplicações (SW)			IMPACTADO	Impacto gerado pela introdução de novos aplicativos	
	Infraestrutura e Segurança			IMPACTADO	Impacto gerado pela introdução de novos aplicativos	
SELEÇÃO E INSTALAÇÃO DE PRODUTOS	Seleção e Instalação de produtos			IMPACTADO	Impacto gerado pela introdução de novos aplicativos	19,20
MODEAGEM DIMENSIONAL	Modelagem Dimensional			IMPACTADO	Impacto gerado pela introdução de novos dados	21,22
PROJETO FÍSICO	Produto Físico			IMPACTADO	Impacto gerado pela introdução de novos dados	23,24
PROJETO E DESENVOLVIMENTO DE ETL	Necessidade do Negócio			IMPACTADO	Impacto gerado pela introdução de novos dados	25,26
	Qualidade de Dados			IMPACTADO	Impacto gerado pela introdução de novos dados	
	Integração de Dados			IMPACTADO	Impacto gerado pela introdução de novos dados	
IMPLANTAÇÃO	Implantação as solução	IMPACTADO	Impacto gerado pela introdução de novos aplicativos e procedimentos e nova forma de identificar requisitos	27,28		
Sem atividades correspondentes ou procedimentos semelhantes que precisam ser avaliadas.		ENTENDER O MODELO DE DADOS	Identificar os requisitos	NOVO	Novos modelos de dados resultantes do processo de KDD	32-42
			Analisar as premissas	NOVO	Novos modelos de dados resultantes do processo de KDD	
			Analisar as restrições	NOVO	Novos modelos de dados resultantes do processo de KDD	
			Avaliar a viabilidade	NOVO	Novos modelos de dados resultantes do processo de KDD	
			Determinar os objetivos	NOVO	Novos modelos de dados resultantes do processo de KDD	
			Identificar os Critérios de Sucesso	NOVO	Novos modelos de dados resultantes do processo de KDD	
			Fazer avaliação inicial da Ferramenta	NOVO/AVALIAR	Novas ferramentas poderão ser inseridas para geração dos modelos de dados resultantes de KDD	
Sem atividades correspondentes ou procedimentos semelhantes que precisam ser avaliadas.		ENTENDER OS DADOS	Coletar os dados adicionais	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	43-49
			Entender os atributos	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
			Verificar as chaves	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
			Avaliar a qualidade dos dados	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
		PREPARAR DADOS	Selecionar Dados	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	50-56
			Limpar Dados	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
			Construir Dados	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
			Construir atributos e atributos derivados	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
			Integrar Dados	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
			Formatar Dados	NOVO/AVALIAR	Esta atividade já pode ter sido realizada para os dados do modelo OLAP	
		CRIAR MODELO DE DADOS	Selecionar a técnica de modelagem	NOVO	Construção do modelo de dados de KDD	57-63
			Definir o teste do Modelo	NOVO	Construção do modelo de dados de KDD	
Construir Modelo	NOVO		Construção do modelo de dados de KDD			
Avaliar tecnicamente o resultado da aplicação do Modelo	NOVO		Construção do modelo de dados de KDD			
AVALIAR RESULTADOS NO NEGÓCIO	Avaliar resultados do Modelo de dados	NOVO	Nova forma de validar a aplicação do Modelo de dados	64-68		
	Revisar as atividades realizadas	NOVO	Avaliar o processo para elaboração do Modelo			
	Determinar próximos passos	NOVO	Introduzir melhorias			
IMPLANTAR A SOLUÇÃO	Implantar o Modelo	NOVO	Levar o modelo para o negócio	69-72		
	Planejar monitoramento e manutenção	NOVO	Validar seu ciclo de vida e necessidade de manutenção			

3.2. DESENVOLVIMENTO DO MÉTODO

O Método foi desenvolvido a partir dos elementos apresentados na Tabela 3, considerando as seguintes premissas definidas pela autora deste trabalho para atendimento dos requisitos definidos para o Método no Capítulo 1 deste trabalho:

- O Método deve fornecer informações necessárias para entender o contexto da sua aplicação.
- O Método deve apresentar seus componentes em forma organizada definindo uma estrutura.
- O Método deve mostrar claramente os impactos da sua aplicação no estado inicial.
- Deve ser identificado o perfil do usuário para o qual o Método está direcionado.
- O Método deve direcionar sua aplicação mediante a execução de ações que, em uma ordem predeterminada, implementam suas atividades.

Nas subseções a seguir estão descritas cada uma das premissas colocadas acima.

3.2.1. Contexto de aplicação do Método

É importante que o Método considere o contexto onde ele é aplicado, trata-se do estado inicial apresentado na Seção 2.4. O conhecimento do contexto de aplicação facilitará a identificação dos impactos na aplicação do Método e avaliação dos recursos computacionais disponíveis e que podem ser aplicados nas ações executadas pelo MÉTODO.

Também deve-se garantir o entendimento do KDD como a nova forma de extrair conhecimento a partir dos dados, este descrito na Seção 2.5.

3.2.2. Estrutura do Método Proposto

O Método está composto por um conjunto de 7 tarefas que se desdobram em suas correspondentes atividades e estas em ações. A Figura 6 apresenta a

composição do Método e seu fluxo de execução. Em particular, são identificados os desvios que representam as possibilidades de iteração do Método. Estes desvios informam o ponto de saída (seta vermelha) para a tarefa e a atividade onde deve ser reiniciada a execução. A atividade ou tarefa onde deve acontecer o reinício do Método também é identificada no bloco azul pelo número que a classifica na estrutura. Quando em lugar do número é colocado “XX”, isto significa que a atividade onde deve ser reiniciado o Método pode variar de com o resultado da execução da atividade em curso. As ações que compõem cada uma das atividades são apresentadas mais adiante nesta seção e estão ilustradas na Figura 6.

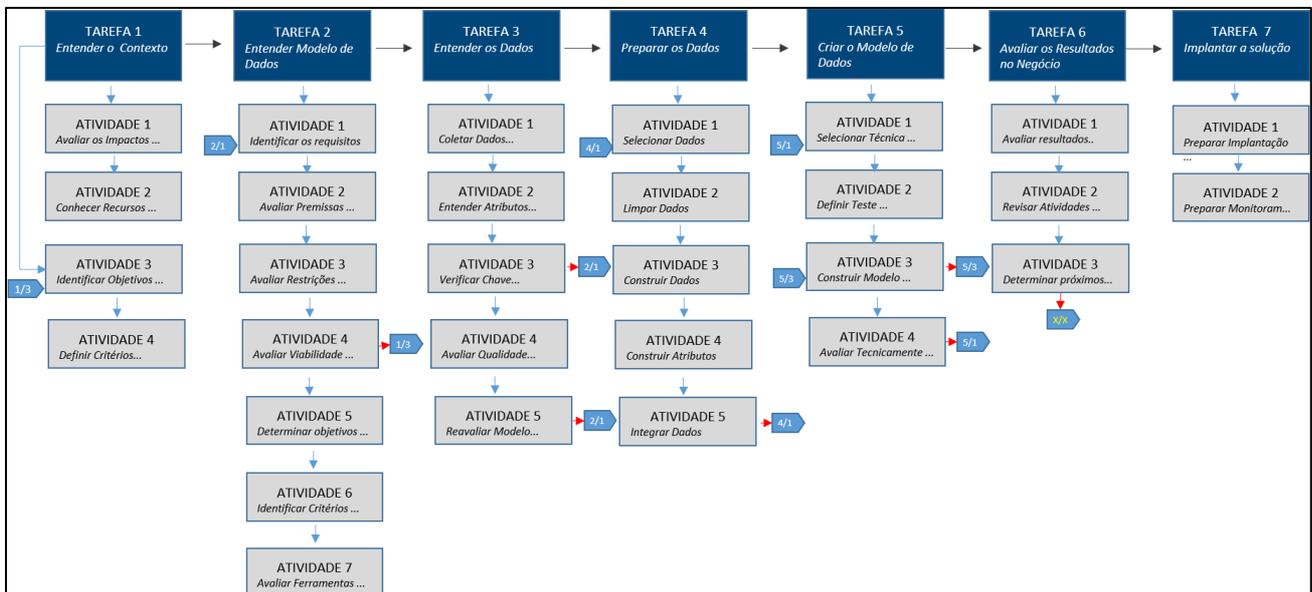


Figura 6 - Estrutura do Método proposto

O caráter iterativo do Método é resultado da execução de algumas atividades. Isto significa que de acordo com uma análise deste resultado, pode ser necessário retornar a tarefas anteriores onde é reiniciada a execução do Método.

3.2.3. Impactos da aplicação do Método

Conforme já colocado na Seção 3.1, a aplicação do Método na plataforma DW/BI pode gerar impactos em alguns componentes da mesma. Estes impactos são gerados pela introdução de 3 novos elementos:

- Novas ferramentas usadas para análise e visualização de dados.

- Novos conjuntos de dados podem fazer parte do *Data Warehouse*.
- Compatibilidade entre procedimentos já executados no estado inicial com procedimentos semelhantes executado no estado final.

A seguir estão descritos cada um dos elementos impactados no estado inicial, conforme indicados na Tabela 3:

Definição dos Requisitos do Negócio. Conforme já colocado na Introdução deste trabalho, em um sistema de BI as hipóteses são levantadas pelo Negócio. Espera-se que os dados irão validar, ou não, estas hipóteses. No processo de KDD, o Negócio levanta questionamentos para os quais os dados informam alguns padrões que posteriormente serão validados pelo Negócio. O correto entendimento de como esta atividade deve ser executada é fundamental, já que o resultado da sua execução se converte em insumos para todas as tarefas que serão executadas posteriormente. Esta afirmativa surge como resultado da análise dos modelos considerados para a estruturar o Método.

Seleção e Instalação dos Produtos. Conforme descrito por Kimball [KIM13], este componente trata do processo de seleção e instalação das ferramentas de software que serão usadas na plataforma DW/BI.

Como resultado da análise do modelo CRISP-DM pode-se afirmar que a seleção inicial de ferramentas para o processo de KDD está associada aos requisitos do Negócio e à capacidade de instalação dentro da arquitetura da plataforma DW/BI. Durante a execução do processo de KDD, os requisitos do Negócio são transformados em requisitos para a formação do Modelo de Dados, fato que introduz novos requisitos para a seleção das ferramentas. Também devem ser consideradas outras variáveis como volume e tipo de dados. Somente em tempo de definição do Modelo de Dados será possível identificar claramente a ferramenta que será usada para a construção do mesmo.

Modelagem Dimensional. A análise da formação do Modelo Dimensional de BI descrita por Kimball permite afirmar que esta atividade deve ser tratada com muito cuidado. Isto se deve ao fato de o procedimento de tratamento realizado como parte desta atividade pode tanto facilitar e apoiar quanto entrar em conflito com o tratamento de dados realizado durante o processo de KDD descrito no modelo CRISP-DM. A decisão de como executar esta atividade dependerá das premissas estabelecidas

pelos requisitos do Negócio em ambos modelos e que gerarão objetivos para a construção do Modelo de Dados. Outro ponto que deve ser considerado como causa de impactos é a introdução de novos dados para atender à demanda do processo de KDD.

Projeto Físico. De acordo com Kimball o Desenho Físico dos dados transforma o Modelo Lógico de dados em uma estrutura física no *Data Warehouse*. O Desenho Físico deve considerar diferentes requisitos como as ferramentas de acesso que serão usadas e a performance dos *queries* realizados sobre os dados. A introdução de novos dados e ferramentas necessários para o processo de KDD pode gerar impactos neste elemento. Cabe ressaltar que a Construção do Modelo Físico surge a partir da definição do Modelo Dimensional, que também pode ter sofrido impactos, e gera requisitos para a elaboração do ETL, que conseqüentemente também poderá se ver impactado.

Projeto e Desenvolvimento do ETL. Conforme Kimball, a construção do ETL está determinada por diferentes requisitos que precisam ser considerados e integrados. Este fato torna esta atividade uma atividade complexa. Entre outros elementos, devem-se considerar: requisitos do negócio, fontes de dados, orçamento e janelas de processamento. De acordo com Kimball, existem 34 subsistemas ou procedimentos que compõem o ETL, fato este que prova a complexidade do processo. Apenas três destes 34 subsistemas poderão sofrer os impactos do Método:

- Necessidade do Negócio: basicamente conduz todo o processo de ETL, lembrando que a necessidade do Negócio é identificada de forma diferenciada para ao processo de KDD.
- Qualidade dos Dados: o tratamento de dados deve estar alinhado com o processo de tratamento de dados, que faz parte do processo de KDD.
- Integração de Dados: novas fontes de dados poderão ser consideradas, inclusive as externas. Portanto, a forma como se fará a integração pode ser afetada.

Implantação. De acordo com Kimball, durante a Implantação, os resultados são apresentados ao Negócio. Considerando a forma diferenciada como os requisitos são identificados, a forma como os resultados são apresentados e validados pelo Negócio também é diferente.

3.2.4. Perfil do Usuário

O público alvo para este Tutorial é o Analista de *BI*. De acordo com Kimball, o Analista de *BI* tem como responsabilidades:

- Garantir que o ambiente de *BI* atenda tecnicamente às necessidades do negócio.
- Desenvolver e fazer manutenção das aplicações de *BI*.
- Configurar as ferramentas de *BI*.
- Fazer a manutenção do banco de dados.
- Entender o negócio e seus dados subjacentes.
- Desenvolver a arquitetura de dados garantindo sua usabilidade, integração e otimização.
- Executar a análise detalhada dos dados para desenvolvimento do Modelo de dados de *BI*.

Além do papel de Analista de BI, Kimball também define um conjunto de papéis de profissionais associados aos aspectos operacionais do seu modelo. Cada um destes profissionais deve ter foco em determinados domínios de conhecimento que lhe darão os subsídios necessários para a execução das atividades associadas ao seu correspondente papel. Para identificar os domínios de conhecimento e o grau de profundidade que o Usuário do Método deve ter em relação aos mesmos, realizou-se uma análise das atividades executadas pelo Método à luz dos domínios de conhecimento identificados para os papeis descritos por Kimball. O nível de profundidade de conhecimento associado a cada domínio foi estabelecido considerando o seguinte critério: a profundidade do conhecimento em domínios de conhecimento que fazem parte do perfil de Analista de BI deve ser “avançado”, os restantes devem ser “básico”. Cabe ressaltar o nível de profundidade básico indica que o Analista de BI deve ter o conhecimento necessário para poder interagir com especialistas dentro dessa área de conhecimento.

A Tabela 4 apresenta a definição do conhecimento necessário para o Usuário do Método fruto da análise acima descrita.

Tabela 4 - Conhecimento necessário para o Usuário do Método

Domínio de Conhecimento	Nível
Técnica para Interação com o Negócio	BÁSICO
Administração de Dados no <i>Data Warehouse</i>	BÁSICO
Arquitetura e desenvolvimento do ETL	BÁSICO
Arquitetura e Infraestrutura da plataforma DW/BI	BÁSICO
Administração dos Metadados na plataforma DW/BI	AVANÇADO
Técnicas e Ferramentas de Mineração de Dados	BÁSICO
Desenvolvimento de Aplicações BI	AVANÇADO
Ferramentas de BI	AVANÇADO
Arquitetura de Dados	AVANÇADO
Preparação e Modelagem de Dados para BI	AVANÇADO

Para o nível de profundidade básico o Usuário do Método precisará de apoio de especialistas para executar as atividades do Método que demandem o uso deste domínio de conhecimento, desta forma o Método se torna totalmente interativo.

Os domínios de conhecimento dentro da organização se traduzem em papéis. Estes papéis podem estar representados pela mesma pessoa ou não, variando de organização para organização. O Método se propõe oferecer as informações necessárias para que o Analista de *BI* se aproprie delas e possa interagir com estes especialistas quando necessário. Lembrando que em alguns casos o próprio Analista de *BI* pode ser ao mesmo tempo especialista em outros domínios de conhecimento. Os papéis aos quais o Método faz referência fazem parte dos papéis descritos por Kimball como necessários para operacionalizar seu modelo:

Analista de Negócios: o Analista de Negócios é responsável por liderar o processo de definição dos requisitos do Negócio e trabalhar como representante destas informações frente à equipe técnica.

- **Administrador de Dados:** o Administrador de Dados é responsável por direcionar a organização na definição de regras do Negócio e domínio de valores permitidos para os dados que compõem o *Data Warehouse*. O Administrador de dados também é responsável pela qualidade de dados carregados no *Data Warehouse*.

- Arquiteto de ETL: o Arquiteto de ETL é responsável pelo processo de extrair, transformar e carregar os dados no Data Warehouse. O Arquiteto de ETL precisa conhecer os sistemas onde os dados se originam e entender o modelo de dados do qual farão parte.
- Arquiteto Técnico: o Arquiteto Técnico é responsável pelo desenho da arquitetura técnica e o suporte técnico para a plataforma DW/BI. Deve garantir que todos os elementos que compõem a arquitetura possam coexistir adequadamente. Tem um papel fundamental na avaliação e seleção dos produtos componentes assim como na sua instalação.
- Gestor de Metadados: o Gestor de Metadados é responsável por todos os metadados que fazem parte da plataforma DW/BI, desde sua identificação e cadastro até a forma como são apresentados para a comunidade do negócio.
- Especialista em Mineração: o Especialista em Mineração é responsável por desenvolver os modelos de dados aplicando metodologias e ferramentas de mineração de dados. Deve ter conhecimento de estatística e do negócio.

3.2.5. Ações que implementam o Método

As Ações que implementam o Método aplicam as considerações já realizadas nesta Seção, a análise das ações que fazem parte do modelo CRISP-DM e outras ações que são resultado da contribuição da autora do Método. Na Tabela 5 apresenta-se a estrutura do Método proposto, onde as atividades são abertas em ações e cada ação é acompanhada de uma justificativa da sua inclusão no Método. Para facilitar a visualização da Tabela esta foi dividida de acordo com as tarefas que a compõem. Para cada ação também é apresentada uma sugestão do papel do especialista que poderá, ou não, interagir com o Usuário do Método, coluna INTERAÇÃO. Dada a complexidade do Método considera-se que o envolvimento dos experts deve gerar resultados mais rápidos e de melhor qualidade. Desta forma, será possível também evitar muitas iterações para completar a execução do Método, estas geradas por falta de conhecimento para a efetiva execução das ações. As ações que geram artefatos de saída estão identificadas com a cor verde. Estes artefatos constam na coluna ENTRADAS, indicando que os mesmos se convertem em entradas para as atividades, fornecendo informações que darão apoio à sua execução. Em azul destacam-se as ações que podem implementar iterações no Método.

Tabela 5 - Ações que implementam o Método proposto

TAREFA 1 - ENTENDER O CONTEXTO				
ATIVIDADE	ENTRADAS	AÇÃO	JUSTIFICATIVA	INTERAÇÃO
1- Avaliar os Impactos da aplicação do Método.		1- Conhecer o contexto de aplicação do Método.	Definida pela Autora com o objetivo de que o usuário conheça os elementos constitutivos da plataforma DW/BI que serão referenciados no Método. Usa como referência o modelo de Kimball.	
		2- Identificar os objetivos do Negócio.	Definida pela Autora com o objetivo de frisar a diferença que existe entre os estados inicial e final para a identificação dos requisitos do Negócio. Usa como referência o modelo de Kimball e o CRISP-DM	Analista de Negócios, Administrador de Dados
		3- Avaliar os possíveis impactos do Projeto de Arquitetura Técnica.	Definida pela Autora com o objetivo de identificar os impactos das aplicação de Método. Usa como referência o modelo de Kimball.	Arquiteto Técnico, Gestor de Metadados, Administrador de Dados
		4- Avaliar os possíveis impactos da seleção e Instalação de Produtos.	Definida pela Autora com o objetivo de identificar os impactos das aplicação de Método. Usa como referência o modelo de Kimball e o modelo CRISP-DM	Analista Técnico, Administrador de Dados
		5- Avaliar os possíveis impactos da Modelagem Dimensional.	Definida pela Autora com o objetivo de identificar os impactos das aplicação de Método. Usa como referência o modelo de Kimball e o modelo CRISP-DM	Analista de Negócios, Administrador de Dados
		6- Avaliar os possíveis impactos da Modelagem Física.	Definida pela Autora com o objetivo de identificar os impactos das aplicação de Método. Usa como referência o modelo de Kimball e o modelo CRISP-DM	Analista de Negócios, Administrador de Dados
		7- Avaliar os possíveis impactos do Desenvolvimento do ETL.	Definida pela Autora com o objetivo de identificar os impactos das aplicação de Método. Usa como referência o modelo de Kimball e o modelo CRISP-DM	Arquiteto de ETL
		8- Avaliar os possíveis impactos da Implantação.	Definida pela Autora com o objetivo de identificar os impactos das aplicação de Método. Usa como referência o modelo de Kimball e o modelo CRISP-DM	Analista de Negócios
		9- Gerar uma RELAÇÃO DE IMPACTOS E CONTINGÊNCIAS.	Definida pela Autora com o objetivo de documentar as informações obtidas como consequência da execução desta atividade. Ter estas informações presentes ajudará a facilitar a execução do Método.	
2- Conhecer os recursos disponíveis.	Relação de Impactos e Contingências	1-Identificar o Conhecimento do usuário do Método.	Definida pela Autora com o objetivo de dar subsídios ao Usuário do Método para obter o conhecimento necessário para a execução. Método. Usa como referência o modelo de Kimball e o modelo CRISP-DM	
		2-Conhecer a plataforma DW/BI.	Definida pela Autora com o objetivo de dar subsídios ao Usuário do Método conhecer a plataforma DW/BI para posterior avaliação dos recursos disponíveis. Usa como referência o modelo de Kimball.	Arquiteto Técnico, Gestor de Metadados, Arquiteto de ETL
		3-Fazer um inventário das Ferramentas de visualização e análise de dados disponíveis.	Definida pela Autora para identificar as ferramentas disponíveis no modelo de partida Vs as potenciais ferramentas que serão usadas no processo de KDD. Existe uma atividades semelhante no modelo CRISP-DM porém não considera o modelo de partida.	Arquiteto Técnico
		4- Avaliar a possibilidade de adquirir novas ferramentas.	Esta ação existe no modelo CRISP-DM.	Arquiteto Técnico
		5- Fazer Inventário dos dados disponíveis.	Definida pela Autora para identificar os dados disponíveis no Data Warehouse. Existe uma atividades semelhante no modelo CRISP-DM porém não considera o modelo de partida.	Administrador de Dados
		6- Verificar a possibilidade de poder obter novos dados de fontes externas.	Esta ação existe no modelo CRISP-DM.	Administrador de Dados
		7- Gerar uma relação de RECURSOS DISPONÍVEIS.	Existe uma ação semelhante no modelo CRISP-DM . O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	
3- Identificar os objetivos do Negócio		1- Identificar a área do Negócio.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Identificar as pessoas chave para mapear os objetivos.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Identificar as áreas do Negócio que são afetadas com os objetivos do KDD.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Definir como identificar os objetivos.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		5- Especificar os benefícios esperados pelo Negócio.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		6- Documentar os OBJETIVOS DO NEGÓCIO	Existe uma ação semelhante no modelo CRISP-DM . O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	
4- Definir os Critérios de sucesso para o Negócio	Objetivos do Negócio	1- Especificar os Critérios de Sucesso para o Negócio (atingíveis e mensuráveis).	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Indicar quem avalia os Critérios de Sucesso.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Incrementar a documentação com os OBJETIVOS DO NEGÓCIO.	Definida pela Autora para dar continuidade ao registro dos Objetivos do Negócio.	

TAREFA 2 - ENTENDER O MODELO DE DADOS				
ATIVIDADE	ENTRADAS	AÇÃO	JUSTIFICATIVA	INTERAÇÃO
1- Identificar os Requisitos para a construção do Modelo de Dados	<i>Objetivos do Negócio, Recursos Disponíveis</i>	1- Identificar o Usuário alvo.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Captar os requisitos do Modelo(Precisão, Capacidade de Manutenção, Repetibilidade).	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Captar outros requisitos adicionais(Segurança, Restrições Legais, Privacidade).	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Gerar uma relação de ESPECIFICAÇÕES DO MODELO DE DADOS	Existe uma ação semelhante no modelo CRISP-DM . O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	
2- Avaliar as premissas para a construção do Modelo de Dados	<i>Recursos Disponíveis, Objetivos do Negócio</i>	1- Esclarecer todas as suposições vinculadas ao conteúdo do conjunto de dados.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Esclarecer todas as suposições em relação aos dados.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios, Analista de Dados
		3- Esclarecer suposições e restrições em relação a fatores externos.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Identificar como os resultados da aplicação do Modelo serão apresentados ao solicitante.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		5- Incrementar a Relação de ESPECIFICAÇÕES DO MODELO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
3- Avaliar as restrições para a construção do Modelo de Dados.	<i>Especificações do Modelo de Dados, Recursos Disponíveis, Objetivos do Negócio</i>	1- Verificar as restrições gerais.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Verificar os direitos de acesso às fontes de dados.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios, Analista de Dados
		3- Verificar acessibilidade técnica dos dados.	Esta ação existe no modelo CRISP-DM.	Analista Técnico
		4- Verificar as restrições orçamentárias.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		5- Incrementar a Relação de ESPECIFICAÇÕES DO MODELO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
4- Avaliar a viabilidade para a construção do Modelo de Dados	<i>Especificações do Modelo de Dados</i>	1- Avaliar a viabilidade de trabalhar com as restrições impostas.	Esta ação existe no modelo CRISP-DM.	
		2- Avaliar se efetivamente as premissas são válidas.	Esta ação existe no modelo CRISP-DM.	
		3- Avaliar a possibilidade de contornar as restrições impostas ou as premissas não válidas.	Esta ação existe no modelo CRISP-DM.	
		4- Se a construção do Modelo de Dados não for viável, volte à TAREFA 1/ATIVIDADE 3 - Identificar os objetivos do Negócio.	Esta ação existe no modelo CRISP-DM. O Autor vinculou a iteração a uma atividade do Método.	
5- Determinar os Objetivos da aplicação do Modelo de Dados	<i>Especificações do Modelo de Dados, Objetivos do Negócio</i>	1- Traduzir os objetivos do Negócio para os objetivos do Modelo de Dados.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Especificar o tipo técnica de modelagem de dados que será usada na construção do Modelo de Dados.	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		3- Incrementar a Relação de ESPECIFICAÇÕES DO MODELO DE DADOS.	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
6- Identificar os Critérios de Sucesso da aplicação do Modelo de Dados	<i>Especificações do Modelo de Dados, Objetivos do Negócio</i>	1- Especificar os critérios técnicos de avaliação do Modelo de Dados	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Definir Benchmarks para os critérios de avaliação.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Identificar possíveis critérios de avaliação subjetivos.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Definir as pessoas que farão a avaliação.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		5- Incrementar a Relação de ESPECIFICAÇÕES DO MODELO DE DADOS.	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
7- Fazer uma avaliação inicial das ferramentas para construir o Modelo de Dados	<i>Especificações do Modelo de Dados</i>	1- Identificar as ferramentas de acordo com as técnicas de modelagem que serão usadas.	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		2- Considerar as ferramentas que já estão disponíveis na sua plataforma DW/BI.	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		3- Escolher ferramentas e técnicas potenciais.	Esta ação existe no modelo CRISP-DM.	Arquiteto Técnico
		4- Validar os requisitos de HW das ferramentas e ver se é possível coloca-las na plataforma DW/BI.	Esta ação existe no modelo CRISP-DM.	Arquiteto Técnico
		5- Gerar uma RELAÇÃO DE FERRAMENTAS contendo as ferramentas potenciais.	Existe uma ação semelhante no modelo CRISP-DM . O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	

TAREFA 3 - ENTENDER OS DADOS				
ATIVIDADE	ENTRADAS	AÇÃO	JUSTIFICATIVA	INTERAÇÃO
1- Coletar os Dados Iniciais	<i>Recursos Disponíveis, Especificações do Modelo de Dados</i>	1- Planejar quais informações são necessárias para elaborar o Modelo de Dados.	Esta ação existe no modelo CRISP-DM.	
		2- Verificar se todas as informações necessárias estão disponíveis.	Esta ação existe no modelo CRISP-DM.	
		3- Especificar os critérios de seleção dos atributos.	Esta ação existe no modelo CRISP-DM.	
		4- Selecionar tabelas e arquivos de interesse.	Esta ação existe no modelo CRISP-DM.	
		5- Selecionar dados dentro de uma tabela ou arquivo.	Esta ação existe no modelo CRISP-DM.	
		6- Considerar quanto tempo um histórico deve ser usado.	Esta ação existe no modelo CRISP-DM.	
		7- Validar a necessidade de obtenção de novos dados de fontes externas.	Esta ação existe no modelo CRISP-DM.	
		8- Gerar um documento de DESCRIÇÃO DE DADOS	Existe uma ação semelhante no modelo CRISP-DM. O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	
2- Entender os atributos	<i>Descrição dos Dados, Especificações do Modelo de Dados, Relação de Impactos e Contingências</i>	1- Verificar acessibilidade e disponibilidade de atributos.	Esta ação existe no modelo CRISP-DM.	
		2- Verificar o tipo de atributos.	Esta ação existe no modelo CRISP-DM.	
		3- Verificar Intervalos de valores dos atributos	Esta ação existe no modelo CRISP-DM. Existe um procedimento no modelo inicial que pode gerar impactos nesta ação.	
		4- Analisar correlações dos atributos.	Esta ação existe no modelo CRISP-DM. Existe um procedimento no modelo inicial que pode gerar impactos nesta ação.	
		5- Entender o significado de cada atributo e o valor do atributo em termos do Negócio.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		6- Para cada atributo calcular estatísticas básicas.	Esta ação existe no modelo CRISP-DM.	
		7- Entrevistar especialistas do domínio para determinar a relevância do atributo.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		8- Incrementar o documento de DESCRIÇÃO DE DADOS.	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
3- Verifique as Chaves	<i>Descrição dos Dados, Especificações do Modelo de Dados, Relação de Impactos e Contingências</i>	1- Analisar os principais relacionamentos	Esta ação existe no modelo CRISP-DM. Existe um procedimento no modelo inicial que pode gerar impactos nesta ação.	
		2- Verificar a quantidade de sobreposições dos principais valores de atributos nas tabelas.	Esta ação existe no modelo CRISP-DM. Existe um procedimento no modelo inicial que pode gerar impactos nesta ação.	
		3- Incrementar o documento de DESCRIÇÃO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
		4- Revisar os objetivos para a construção do Modelo de Dados.	Esta ação existe no modelo CRISP-DM.	
		5- Se os objetivos do Modelo de Dados não forem viáveis volte à TRFEA2/ATIVIDADE 1 - Identificar os Requisitos para a Construção do Modelo de Dados	Existe uma ação semelhante no modelo CRISP-DM. O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	
4- Avaliar a qualidade dos Dados	<i>Descrição dos Dados, Relação de Impactos e Contingências</i>	1- Considerar a revisão de chaves e atributos.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	
		2- Verificar a cobertura.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Analista de Negócios
		3- Verificar se o significado dos atributos e valores contidos se encaixam.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Analista de Negócios
		4- Identificar atributos ausentes e campos em branco.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	
		5- Estabelecer o significado de dados ausentes.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Analista de Negócios
		6- Verificar se há atributos com valores diferentes e com significados semelhantes.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Analista de Negócios
		7- verificar a ortografia e o formato dos valores.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	
		8- Verificar se há ruídos.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Analista de Negócios
		9- Verificar a plausibilidade dos valores.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		10- Incrementar o documento DESCRIÇÃO DE DADOS.	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
5- Reavaliar o Modelo e formatar suposições para análise futura	<i>Descrição dos Dados, Especificações do Modelo de Dados</i>	1- Formular questões para avaliar o Modelo de Dados.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Transformar as questões em possíveis resultados.	Esta ação existe no modelo CRISP-DM.	
		3- Realizar uma análise básica para avaliar estes resultados.	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Esclarecer ou tornar mais precisos os objetivos Modelo de Dados seus critérios de sucesso.	Esta ação existe no modelo CRISP-DM.	
		5- Incrementar a ESPECIFICAÇÃO DO MODELO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
		6- Se os resultados desta análise mostrar a necessidade de fazer mudanças no Modelo de Dados e não forem viáveis, volte à TAREFA 2/ATIVIDADE 1 - Identificar os requisitos para a construção do Modelo de Dados.	Existe uma ação semelhante no modelo CRISP-DM. O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	

TAREFA 4 - PREPARAR OS DADOS				
ATIVIDADE	ENTRADAS	AÇÃO	JUSTIFICATIVA	INTERAÇÃO
1- Selecionar os Dados	<i>Especificações do Modelo de Dados, Descrição do Dados</i>	1- Coletar dados de diferentes fontes.	Esta ação existe no modelo CRISP-DM.	Administrador de Dados
		2- Considerar a seleção de subconjuntos de Dados diferentes.	Esta ação existe no modelo CRISP-DM.	
		3- Considerar o uso de técnicas de amostragem.	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		4- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS.	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
2- Limpar Dados	<i>Especificações do Modelo de Dados, Descrição do Dados, Relação de Impactos e Contingências</i>	1- Considerar como lidar com qualquer tipo de ruído observado.	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Administrador de Dados
		2- Corrigir, remover ou ignorar ruído	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Administrador de Dados
		3- Decidir como lidar com valores especiais e seus significados	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Administrador de Dados
		4-Verificar mecanismos de construção disponíveis com a lista de ferramentas identificadas	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	
		5- Decidir se é melhor realizar a construção dentro da ferramenta ou fora dela	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	
		6- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
3- Construir Dados	<i>Especificações do Modelo de Dados, Descrição do Dados</i>	1- Verificar os mecanismos de construção disponíveis com a Lista de Ferramentas identificadas	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		2- Decidir se é melhor realizar a construção dentro ou fora da ferramenta.	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		3- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
4- Construir atributos e atributos derivados	<i>Especificações do Modelo de Dados, Descrição de Dados, Relação de Ferramenta, Relação de Impactos e Contingências</i>	1- Considerar os seguintes critérios para a construção de atributos.(O conhecimento de fundo nos convence de que algum fato é importante e deve ser representado, embora ao tenhamos um atributo atual para representá-lo/A algoritmo de modelagem em uso lida apenas com certos tipos de dados/A avaliação do resultado esperado do Modelo sugere que certos fatos não estão sendo cobertos/Adicionar novas informações sobre a importância relevante dos atributos, adicionando novos atributos/Construir ou imputar atributos ausentes/Adicionar novos atributos aos dados acessados/transformar atributos	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Especialista em Mineração
5- Integrar Dados	<i>Especificações do Modelo de Dados, Descrição do Dados, Relação de Ferramentas</i>	1- Verificar se as ferramentas de integração são capazes de integrar as fontes de entrada conforme necessária	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	Especialista em Mineração
		2- Integrar fontes e armazenar resultados	Esta ação existe no modelo CRISP-DM. Existe um procedimento semelhante no modelo inicial que pode gerar impactos nesta ação.	
		3- Reconsiderar Critérios de Seleção de Dados	Esta ação existe no modelo CRISP-DM.	
		4- Se os resultados desta análise mostrar a necessidade de fazer mudanças no Modelo de Dados e estas não forem viáveis voltar a TAREFA 4/ATIVIDADE 1 - Selecionar Dados	Não existe a necessidade de fazer mudanças no Modelo de Dados e portanto de fazer uma iteração na execução do Método	

TAREFA 5 - CRIAR O MODELO DE DADOS				
ATIVIDADE	ENTRADAS	AÇÃO	JUSTIFICATIVA	INTERAÇÃO
1- Selecionar a Técnica de Modelagem	<i>Especificações do Modelo de Dados, Descrição do Dados, Relação de Ferramentas</i>	1- Considerar todo o conhecimento já adquirido sobre os dados	Esta ação existe no modelo CRISP-DM.	
		2- Considerar os resultados esperados	Esta ação existe no modelo CRISP-DM.	
		3- Considerar a Ferramenta	Esta ação existe no modelo CRISP-DM.	
		4- Selecionar a técnica de modelagem	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		5- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
2- Definir o teste do Modelo de Dados	<i>Especificações do Modelo de Dados, Descrição do Dados</i>	1- Verificar diferentes testes para cada uma das metas definidas para o Modelo de Dados	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		2- Definir os passos necessários	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		3- Preparar os dados necessários para o teste	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		4- Selecionar a técnica de modelagem	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		5- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	
3- Construir o Modelo	<i>Especificações do Modelo de Dados, Descrição do Dados, Relação de Ferramentas</i>	1- Definir parâmetros iniciais da ferramenta	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		2- Executar a técnica selecionada no conjunto de dados de entrada para produzir o Modelo de Dados	Esta ação existe no modelo CRISP-DM.	Especialista em Mineração
		3- Avaliar o resultado	Esta ação existe no modelo CRISP-DM.	
		4- Interpretar o resultado do Modelo	Esta ação existe no modelo CRISP-DM.	
		5-- Se os resultados não são satisfatórios construa outro Modelos e compare os resultados. Com este objetivo volte a TAREFAS/Atividade 3 - Construir o Modelo	Não existe a necessidade de fazer mudanças no Modelo de Dados e portanto de fazer uma iteração na execução do Método	
4- Avaliar tecnicamente o Modelo de Dados	<i>Especificações do Modelo de Dados, Descrição do Dados, Objetivos do Negócio</i>	1- Avaliar os resultados em relação aos critérios de avaliação do Modelo de Dados	Esta ação existe no modelo CRISP-DM.	
		2- Criar ranking de resultados com respeito a critérios de sucesso e avaliação	Esta ação existe no modelo CRISP-DM.	
		3- Interpretar os resultados em termos do Negócio - tanto quanto possível nesta fase	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Obter comentários dos especialistas do domínio sobre estes resultados	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		5- Verificar a plausibilidade do Método	Esta ação existe no modelo CRISP-DM.	
		6- Se os resultados desta avaliação não atenderem os critérios estabelecidos volte à TAREFA 5- Criar Modelo de Dados	Não existe a necessidade de fazer mudanças no Modelo de Dados e portanto de fazer uma iteração na execução do Método	

TAREFA 6 - AVALIAR OS RESULTADOS NO NEGÓCIO				
ATIVIDADE	ENTRADAS	AÇÃO	JUSTIFICATIVA	INTERAÇÃO
1- Avaliar o resultado do Modelo de Dados	<i>Objetivos do Negócio, Modelo de Dados</i>	1- Compreender e interpretar os resultados apresentados pelo Modelo de Dados	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Verificar se conhecimento contido e novo e útil para o Negócio	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Avaliar os resultados em relação aos critérios de sucesso do Negócio	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
2- Revisar Tarefas	<i>Relação de Recursos Disponíveis, Descrição de Dados, Relação de Impactos e Contingências, Especificações do Modelo de Dados, Relação de Ferramentas, Descrição de Dados, Objetivos do Negócio</i>	1- Fazer uma avaliação geral em função dos resultados obtidos	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Avaliar cada Atividade realizada questionando: Foi necessária?/Foi executada de forma otimizada?/De que maneiras poderia ser melhorada?	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Identificar falhas	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- identificar possíveis ações alternativas e/ou caminhos inesperados	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
3- Determinar próximos passos	<i>Relação de Recursos Disponíveis, Descrição de Dados, Relação de Impactos e Contingências, Especificações do Modelo de Dados, Relação de Ferramentas, Descrição de Dados, Objetivos do Negócio</i>	1- Analisar o potencial de implantação do resultado	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Estimar o potencial de melhoria do processo atual	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Avaliar opções alternativas de continuidade	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Reiniciar o processo na tarefa de acordo com a avaliação anterior . TAREFA XX/ Atividade XX	Após a avaliação toda a execução do Método, se identificada a necessidade volta-se a uma /atividade anterior de acordo com o resultado da avaliação. "XX" indica um número genérico.	

TAREFA 7 - IMPLANTAR A SOLUÇÃO				
ATIVIDADE	ENTRADAS	AÇÃO	JUSTIFICATIVA	INTERAÇÃO
1- Planejar a implantação	<i>Objetivos do Negócio</i>	1- Determinar como o conhecimento ou a informação será propagada	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Decidir como o uso do resultado será monitorado e seus benefícios medidos	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Desenvolver e gerar um PLANO DE IMPLANTAÇÃO	Existe uma ação semelhante no modelo CRISP-DM. O Autor definiu um novo layout para este relatório devido às mudanças realizadas nas ações que geram as informações que constam no relatório.	
2- Planejar Monitoramento e Manutenção	<i>Objetivos do Negócio, Plano de Implantação</i>	1- Verificar se há aspectos dinâmicos ou temporais no Negócio que afetem o resultado do Modelo	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		2- Definir uma forma (plano) de Monitoramento	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		3- Determinar quando o resultado ou o Modelo de dados não deve ser mais usado	Esta ação existe no modelo CRISP-DM.	Analista de Negócios
		4- Incrementar o PLANO DE IMPLANTAÇÃO	Definida pela Autora para dar continuidade ao registro das Especificações do Modelo de Dados.	

3.3. CONCLUSÕES

O Método proposto neste Capítulo é aplicado sobre uma plataforma DW/BI onde é avaliado o impacto da sua aplicação visando a manutenção das funcionalidades da estrutura. Os requisitos para iniciar a execução do Método tem como origem o Negócio. Quanto aos resultados, novos padrões de dados, são entregues ao Negócio para sua avaliação. O Método é implementado por um conjunto de Tarefas que por sua vez se desdobram em Atividades e Ações que possibilitam sua execução em forma organizada. Estas ações focam apenas em questões computacionais: identificação de requisitos, seleção e tratamento de dados, construção do Modelo de Dados e sua validação técnica, entrega dos resultados e finalmente as ferramentas necessárias para apoio à execução das ações. Face ao exposto, pode-se afirmar que o Método desenvolvido atende aos objetivos e requisitos inicialmente propostos para este trabalho. Entretanto, é relevante avaliar-se esta conclusão com especialistas.

4. AVALIAÇÃO DO MÉTODO

Este Capítulo apresenta os mecanismos utilizados para avaliação do Método e os resultados obtidos. A Seção 4.1. apresenta a modelo de avaliação TAM (*Technology Acceptance Model*). A Seção 4.2 apresenta os Grupos Focais. A Seção 4.3 apresenta o Tutorial. A Seção 4.4 apresenta a avaliação propriamente dita, os resultados obtidos e a conclusões.

4.1. APLICAÇÃO DOS CRITÉRIOS DE UTILIDADE E FACILIDADE DE USO DO MODELO DE AVALIAÇÃO TAM

De acordo com March et al. [MAR95], a tecnologia pode ser definida como implementações práticas de inteligência. March considera a tecnologia como uma prática útil, em vez de ser um fim em si mesmo, fornecendo ferramentas e técnicas desenvolvidas pelos seres humanos para alcançar objetivos. O autor define um Método como um conjunto de etapas que levam à execução de uma tarefa. Os Métodos são frequentemente usados para traduzir um modelo ou representação para outro novo modelo, no curso de resolver um problema. March, acrescenta que as pesquisas em TI são instanciadas, se operacionalizam, mediante artefatos: Construções, Métodos e Modelos. As instanciações demonstram a viabilidade e eficácia dos Modelos e Métodos. Assim, os artefatos são construídos para executar uma tarefa específica. A questão primordial refere-se ao seu funcionamento. Desta forma, estes artefatos tornam-se o objeto de estudo. Na sua avaliação, identifica-se se com o Método proposto obteve-se algum progresso; ou seja, se a nova tecnologia é mais eficaz que aquela que se deseja substituir.

Para fazer a avaliação do Método proposto foi criado um artefato, um Tutorial. O Tutorial está composto por um conjunto de instruções que conduzem, passo a passo, à implementação do Método. Desta forma a nova tecnologia pode ser efetivamente instanciada.

O Tutorial foi avaliado com a aplicação de um questionário construído a partir dos critérios de avaliação de Utilidade e Facilidade de Uso que fazem parte do Modelo de Aceitação de Tecnologia (TAM – *Technology Acceptance Model*) [DAV89]. De acordo com Davis [DAV89], a razão para usar estes indicadores, utilidade e facilidade

de uso, é que eles estão fortemente correlacionados com a aceitação do usuário de uma determinada tecnologia. Estes indicadores são avaliados mediante a aplicação de um formulário onde os usuários respondem a perguntas sobre o objeto em avaliação usando uma pontuação. Para o estudo da aplicação prática desta avaliação do modelo TAM, alguns trabalhos foram considerados: [FRE13], [MAE06] e [RIV15].

A Tabela 6 apresenta os critérios para avaliação da Utilidade propostos e a Tabela 7 apresenta os critérios para avaliação da Facilidade de Uso do mesmo autor.

Tabela 6 - Critérios de Utilidade do modelo TAM [DAV89]

	Item
1	Dificuldades no trabalho
2	Controle no trabalho
3	Desempenho no trabalho
4	Abordagem das necessidades do trabalho
5	Economiza tempo no trabalho
6	Rapidez no trabalho
7	Decisivo no trabalho
8	Completar mais tarefas no trabalho
9	Reduzir o tempo improdutivo no trabalho
10	Eficácia no trabalho
11	Qualidade do trabalho
12	Aumento da Produtividade do trabalho
13	Torna o trabalho mais fácil
14	É útil no trabalho

Tabela 7 - Critérios de Facilidade de uso do modelo TAM [DAV89]

	Item
1	Confuso
2	Erros frequentes
3	Frustrante
4	Dependência do manual
5	Esforço mental
6	Reparação de erros
7	Rígido e inflexível
8	Controlável
9	Comportamento inesperado
10	Cansativo
11	De difícil compreensão
12	Facilidade de lembrar
13	Fornecer orientações úteis
14	Fácil de usar
15	Facilidade de Aprendizagem
16	Facilidade de Percepção

Algumas adaptações foram feitas nas tabelas originais para obter um melhor alinhamento com o Método que será avaliado. As Tabelas 8 e 9 apresentam e justificam as adaptações realizadas na Tablas de avaliação de Utilidade e de Facilidade de uso respectivamente. No Apêndice C constam os questionários construídos a partir destas tabelas.

Tabela 8 - Adaptações realizadas pela autora na Tabela de Utilidade do modelo TAM

	ITEM	QUESTÃO	MOTIVO
1	Dificuldade no trabalho	A	
2	Controle no trabalho	B	
3	Desempenho no trabalho	C	
4	Abordagem das necessidades do trabalho	D	
5	Economiza tempo no trabalho	E	
6	Rapidez no trabalho	F	
7	Decisivo no trabalho	G	
8	Completar mais tarefas no trabalho	H	
9	Reduzir o tempo improdutivo no trabalho	NÃO SE APLICA	Não é possível verificar e eficácia a produtividade de um processo novo
10	Eficácia no trabalho	NÃO SE APLICA	Não é possível verificar e eficácia de um processo novo
11	Qualidade do trabalho	NÃO SE APLICA	Não é possível verificar a qualidade de um processo novo
12	Aumento da Produtividade do trabalho	NÃO SE APLICA	aumento de produtividade na aplicação de um processo novo.
13	Torna o trabalho mais fácil	I	
14	É útil no trabalho	J	

Tabela 9 - Adaptações realizadas pela autora na Tabela de Facilidade do modelo TAM

	ITEM	QUESTÃO	MOTIVO
1	Confuso	A	
2	Erros freqüentes	B	
3	Frustrante	C	
4	Dependência do manual	NÃO SE APLICA	Não existe um manual sobre o Método
5	Esforço mental	D	
6	Reparação de erros	NÃO SE APLICA	O Método não oferece a oportunidade de reparar erros
7	Rígido e inflexível	E	
8	Controlável	F	
9	Comportamento inesperado	G	
10	Cansativo	H	
11	De difícil compreensão	I	
12	Facilidade de lembrar	J	
13	Fornecer orientações úteis	K	
14	Fácil de usar	L	
15	Facilidade de Aprendizagem	NÃO SE APLICA	O Método não objetiva uma aprendizagem
16	Facilidade de Percepção	NÃO SE APLICA	Esta característica não se aplica à usabilidade do Método

Como forma de apontar o nível de concordância ou não com as questões realizadas, Davis [DAV89] propõe 4 níveis de pontuação: extremo, bastante, levemente e indiferente. Este formato é pouco usado e considera-se que pode apresentar dúvidas na sua aplicação. De acordo com a recomendação de Hinojosa [HIN07] foi escolhida a escala Likert com 5 opções: Nenhum, Pouco, Regular, Bom e Muito Bom. As questões foram formuladas de forma de possibilitar e estimular os respondentes a acessar a primeira e última opção.

4.2. GRUPOS FOCAIS

Morgan [MOR96] define grupos focais como uma técnica de pesquisa qualitativa, derivada das entrevistas grupais, que coleta informações por meio das interações entre os integrantes de um grupo. A utilização de grupos focais, de forma isolada ou combinada com outras técnicas de coleta de dados primários, revela-se especialmente útil na pesquisa avaliativa. Seu principal objetivo é reunir informações detalhadas sobre um tópico específico (sugerido por um pesquisador, coordenador ou moderador do grupo) a partir de um grupo de participantes selecionados.

4.3. O TUTORIAL

O Tutorial descreve o conjunto de ações que implementam o Método. Ele foi construído usando *slides* que aplicam diferentes recursos visuais para facilitar seu uso. Os *slides* são acompanhados de relatos, som, que explicam o conteúdo de cada *slide* e fornecem exemplos que ajudam no entendimento do conteúdo. *Slides* contendo informações adicionais e outras figuras foram adicionadas pela autora com o objetivo de tornar o Tutorial mais didático. O Tutorial inicia sua execução instruindo o usuário sobre seu funcionamento e estrutura. Logo a seguir é apresentada a implementação do Método.

No Apêndice A constam os *slides* que compõem o Tutorial, acompanhados dos scripts com o relato explicativo sobre seu conteúdo.

A maior parte do material usado para construir o Tutorial encontra-se referenciado neste trabalho. O restante são contribuições do autor ou imagens extraídas de *sites*, suas respectivas fontes encontram-se referenciadas no *Slide* onde o material é exibido.

4.4. EXECUÇÃO DA AVALIAÇÃO

A avaliação do Método foi executada em duas etapas. Uma avaliação inicial foi realizada por especialistas. Posteriormente, uma segunda avaliação foi realizada pelo grupo focal também formado por especialistas.

4.4.1. Avaliação inicial

A avaliação inicial teve como objetivo constatar se o material apresentado podia considerar-se apto para apresentação ao grupo focal. Os dois especialistas que realizaram esta primeira avaliação têm o seguinte perfil: profissionais experientes do mercado com domínio de conhecimento tanto em BI quanto em Ciência de Dados. Ambos com formação MSc. em Tecnologia da Informação, com foco em Ciência de Dados.

Para esta avaliação inicial foi apresentado o Tutorial e posteriormente o questionário para avaliação. Como resultado desta avaliação inicial obteve-se o seguinte retorno:

Especialista 1: O Método atende a uma necessidade que muitas organizações vivenciam hoje. O Tutorial é muito longo (duração mínima de uma hora), porém consegue descrever passo a passo as ações necessárias para cumprir seus objetivos. Algumas das terminologias usadas são muito “acadêmicas”, diferentes das usadas nas práticas de mercado. O questionário de avaliação foi considerado adequado.

Especialista 2: O Método é muito extenso e cansativo. Usaria o Tutorial como uma forma de treinar a equipe da sua organização, com algumas restrições oriundas da observação anterior. Algumas das questões apresentadas para a avaliação do Método devem ser reformuladas para facilitar seu entendimento.

Como resultado das observações realizadas pelos especialistas foram executadas as seguintes ações:

- O Método é complexo e extenso. O Tutorial é muito longo. A transformação que o Método implementa é realmente complexa e isto se evidencia no número de ações que compõem o Método, aproximadamente 200, e no tempo de duração do Tutorial. Em relação a esta observação nenhuma ação foi realizada.

- Uso de terminologia considerada “acadêmica”. Em relação a esta observação os termos assinalados foram avaliados, chegando-se à conclusão de que sua aplicação é correta e adequada. Verificou-se o Tutorial, inserindo a definição destes conceitos no seu conteúdo.

- Algumas das questões apresentadas nos questionários para a avaliação do Método devem ser reformuladas para facilitar seu entendimento. Em relação a esta observação, as questões assinaladas foram avaliadas e modificadas.

4.4.2. Avaliação do Grupo Focal

Para melhor organizar as informações esta subseção foi estruturada da seguinte forma: inicia com a descrição da etapa Preparação, esta descreve as ações realizadas para a preparação da reunião do grupo focal. A seguir descreve-se a etapa Execução, relatando como aconteceu a reunião do grupo focal. Finalizando a etapa Resultados apresenta os resultados obtidos e as correspondentes considerações da autora.

4.4.2.1. Preparação

O trabalho de Barbour[BAR09] foi usado como guia para a preparação da reunião com o grupo focal. Este trabalho apresenta uma análise profunda dos detalhes da dinâmica de trabalho com grupos focais, abordando a preparação do encontro e ressaltando muitas situações corriqueiras e como enfrenta-las. Assim sendo, foram considerados os seguintes aspectos para preparar a reunião do grupo focal:

- Escolha do público-alvo.
- Motivação para a participação.
- Preparação do ambiente para o encontro.
- Previsão de dificuldades potenciais.
- Preparação do material apresentado ao grupo.
- Preparação de materiais de estímulo.
- Recolhimento e registro de informações.
- Pós-participação.

O público-alvo foi definido de acordo ao usuário ao qual está destinado o Método. Doze Analistas de BI, selecionado por conveniência, foram convidados por *e-mail* para participar do encontro. Solicitou-se confirmação da participação, contando-se com o retorno de 10 especialistas. O *e-mail* enviado fazia um breve relato do objetivo do encontro, frisando na importância que a transição implementada pelo Método tem no dia de hoje para as organizações. Acompanhando o *e-mail* foram enviadas informações adicionais apresentando os melhores resultados obtidos por organizações que já fazem uso do KDD.

Diversas ações foram tomadas para preparar a reunião do grupo tendo presente a previsão de algumas dificuldades:

- Adicionar ao Tutorial um filme, com música, para fazer uma apresentação inicial de forma que chamasse a atenção dos participantes.
- Preparar um questionário de abertura para avaliação do nível de conhecimento dos participantes em relação aos domínios de conhecimento presentes na definição do Método. Os domínios de conhecimento considerados em este questionário foram extraídos da Tabela 4 que consta no Capítulo 3 deste trabalho. O

nível de profundidade de cada domínio de conhecimento avaliado usando a escala Likert, apresentando as seguintes opções: Nenhum, Pouco, Regular, Bom e Muito Bom. O questionário é apresentado no Apêndice C.

- Adicionar ao questionário de avaliação a solicitação de informações de identificação do participante. Buscou-se identificar a senioridade do respondente mediante a identificação da função, tempo da empresa e tempo na função.

- Preparar uma pessoa para auxiliar na condução do encontro, com as seguintes responsabilidades: passar os *slides* do Tutorial, acompanhar a gravação da discussão do grupo sobre a apresentação, fazer anotações sobre os pontos mais relevantes discutidos pelo grupo.

- Preparar um questionário motivacional para iniciar as discussões após a apresentação do Tutorial.

- Enviar mais um *e-mail* no dia anterior ao encontro buscando garantir a presença dos participantes.

4.4.2.2. Execução

No dia do encontro do grupo focal oito pessoas compareceram. Foi oferecido um lanche para fazer a recepção e ao mesmo tempo criar um ambiente mais descontraído para a reunião. Antes de iniciar a apresentação, solicitou-se aos participantes o preenchimento do questionamento de avaliação dos Domínios de Conhecimento. Também foi entregue material, papel e caneta, para que os participantes pudessem registrar seu parecer durante a apresentação do Tutorial, já que esta não seria interrompida. A apresentação do Tutorial demorou uma hora. Seguidamente solicitou-se aos participantes preencher os questionários referentes à avaliação TAM. Logo após começaram as discussões entre os participantes. No papel de mediador, a autora deste trabalho que conta com experiência previa na condução de *workshops*. Não foi preciso motivar a discussão, esta fluiu facilmente. As discussões se prolongaram por aproximadamente 55 minutos. Já ultrapassado o tempo planejado, e dada o avançada da hora, os participantes começaram a sair e desta forma foram encerradas as atividades.

4.4.2.3. Resultados Obtidos

De acordo com a identificação dos Domínios de Conhecimento, os participantes mostraram o conhecimento adequado para avaliar o conteúdo do Tutorial. A Tabela 10 apresenta o resultado da avaliação. Em cada célula é apresentado o número de respondentes para a questão (linha) de acordo com a classificação (1/Nenhum, 2/Pouco, 3/Regular, 4/Bom, 5/Muito Bom). Na última coluna, ESPERADO, acrescentou-se o nível de conhecimento esperado sobre cada domínio de conhecimento para o usuário do Tutorial de acordo com o Capítulo 4 deste trabalho. Um dos presentes não entregou o questionário.

Tabela 10 – Avaliação dos Domínios de Conhecimento dos participantes do Grupo Focal.

DOMÍNIOS de CONHECIMENTO							
		1	2	3	4	5	ESPERADO
A	Técnicas para Interação com o Negócio			1	4	2	BÁSICO
B	Administração dos dados no Data Warehouse		1	2	2	2	BÁSICO
C	Arquitetura e desenvolvimento do ETL		2	2	1	2	BÁSICO
E	Arquitetura e Infraestrutura da plataforma DW/BI		3	1	2	1	BÁSICO
F	Administração dos Metadados na Plataforma DW/BI		2	3	1	2	AVANÇADO
G	Técnicas e Ferramentas de Mineração de Dados	1	2	3		1	BÁSICO
H	Desenvolvimento de aplicações BI			1	3	3	AVANÇADO
I	Ferramentas BI		1	1	2	3	AVANÇADO
J	Arquitetura de Dados		1	2	1	3	AVANÇADO
K	Preparação e Modelagem de Dados para BI		1	2	1	3	AVANÇADO

A identificação da função desempenhada pelos participantes mostrou os seguintes resultados:

- Seis dos presentes identificaram-se como Analistas de BI.
- Um dos presentes identificou-se como Gerente de BI (função que ele mesmo acrescentou à lista de funções apresentadas).
- Um dos presentes identificou-se como Analista de Dados.
- Dois dos presentes já identificados como Analistas de BI acrescentaram mais funções: um deles acrescentou a função de Arquiteto de ETL e de Analista de Negócios. O segundo acrescentou a função Analista de Negócios.

Em relação ao tempo na função os valores apresentados variaram entre 2 e 30 anos, sendo que um dos participantes não respondeu a este questionário.

Na avaliação das informações apresentadas, conclui-se que o perfil dos avaliadores se encontra dentro do perfil do Usuário do Método. O nível de conhecimento é adequado e está de acordo com os Domínios de Conhecimento e profundidade necessária identificados, no Capítulo 3 deste trabalho, para aplicação do Método. Os avaliadores reconhecem-se fazendo parte da função identificada para a aplicação do Método e a senioridade dos mesmos constatou-se como de nível médio.

O resultado da aplicação dos questionários usados para avaliar o Tutorial apresentam-se nas Tabela 11 e na Tabela 12.

Tabela 11 - resultado da avaliação da Utilidade do Método aplicando TAM

UTILIDADE		1	2	3	4	5
A	O Método simplifica muito as tarefas.			5	2	1
B	O Método direciona claramente as tarefas.				7	1
C	O Método aumenta muito o desempenho.			4	2	2
D	O Método endereça claramente seu objetivo.			1	6	1
E	O Método economiza muito tempo.		1	1	5	1
F	O Método permite executar as tarefas com rapidez.		1	3	3	1
G	O Método endereça aspectos muito críticos.		1	1	4	2
H	O Método é fundamental para conseguir completar as tarefas.			3	3	2
I	O Método é um grande facilitador.			2	4	2
J	O Método é extremamente útil.			2	4	2
			3	22	40	15

Tabela 12 - Resultado da avaliação da Facilidade de Uso do Método aplicando TAM

FACILIDADE DE USO		1	2	3	4	5
A	O Método não gera nenhum tipo de confusão.		2	3	2	1
B	O Método não provoca nenhum tipo de erro.		3	3	1	1
C	O Método não me decepciona em forma alguma.		2	2	3	1
D	O Método não demanda um grande esforço mental.		4	2	2	
E	O Método não é, em forma alguma, rígido ou inflexível.		1	2	4	1
F	O Método é muito claro.		1	2	4	1
G	O Método não gera situações inesperadas.		1	5	2	
H	O Método não é cansativo em forma alguma.		3	4	1	
I	O Método é de muito fácil entendimento.		2	5		1
J	O Método é muito fácil de ser aprendido.		2	5		1
K	O Método me fornece orientações muito úteis.		1	1	4	2
L	O Método é, acima de tudo, muito fácil de usar.		2	5		1
			24	39	23	10

Na análise da Tabela 11, pode-se observar uma forte tendência das respostas a indicar o Método como Útil. Destacam-se as seguintes questões com avaliação mais positiva: direcionamento claro das tarefas, endereçamento claro do objetivo,

possibilita economia de tempo, endereçamento de aspectos críticos e de utilidade. Destacam-se as seguintes questões com avaliação menos positiva: simplifica a execução das tarefas, aumento do desempenho.

Na análise da Tabela 12, pode-se observar uma tendência a uma imparcialidade em relação à Facilidade de Uso. Destacam-se as seguintes questões com avaliação mais positiva: não gera decepção, não é rígido ou inflexível, é claro, oferece orientações muito úteis. Destacam-se as seguintes questões com avaliação mais negativa: não é fácil de usar, não é fácil de ser aprendido, é cansativo, demanda um grande esforço mental.

Finalizando, segue as colocações realizadas pelos participantes após a apresentação do Tutorial juntamente com o resultado das anotações recebidas.

- Todos concordaram em que o Tutorial é muito demorado.
- Todos concordaram com que o Método aborda um tema de alta relevância para a comunidade de Analistas de BI.
- Todos concordaram com que o Método direciona claramente, passo a passo, ao seu objetivo.
- Uma parte dos presentes concordou que as orientações oferecidas pelo Método são úteis, especialmente reforçadas pelas informações apresentadas nas tabelas que fazem parte do Tutorial.
- Grande parte dos participantes colocaram que consideram o Método muito complexo elencando muitas atividades. Ressaltou-se que por este motivo não fizeram as anotações solicitadas já que a apresentação demandava o máximo de atenção.
- Como solução à alta complexidade do Método, um dos participantes mencionou o uso de uma metodologia semelhante à metodologia Scrum. Explicando melhor a ideia entendeu-se que em realidade ele sugeria uma construção incremental. O grupo depurou esta ideia e elaborou a seguinte proposta: o Tutorial devia ser apresentado em camadas. Cada camada apresentaria um nível mais detalhado na descrição das ações, ficando por conta de o usuário decidir aprofundar nas camadas ou não. Com esta estrutura, as camadas superiores seriam mais simples e o nível de complexidade aumentaria na medida que se aprofunda nas camadas.

- As tabelas com informações adicionais apresentadas no Tutorial foram consideradas de grande utilidade. Nestas tabelas informações de grande valor encontram-se condensadas e são de fácil entendimento. Chegou-se a um consenso que estas tabelas deveriam ser mais exploradas e enriquecidas.
- Os exemplos colocados na execução das atividades foram considerados muito úteis, porém escassos. Sugeriu-se colocar exemplos em todas as atividades do Tutorial.
- Considerou-se interessante aplicar o Tutorial como forma de treinamento do Método em organizações que pretendem fazer a transição, porém ele deveria ser simplificado ou particionado e incrementado de exercícios.
- Foi discutida a utilidade das ferramentas de código aberto e a necessidade de inserir outras na relação apresentada pelo Tutorial.
- Alguns participantes colocaram que a indicação de especialistas é os pontos onde eles devem ser chamados, é totalmente pertinente dada a complexidade do Método e a tangência com outros domínios de Conhecimento.

Deve-se ressaltar que os tópicos apresentados pelos participantes se mostraram convergentes. Poucas discussões foram realizadas sobre cada um dos pontos colocados já que houve concordância entre os participantes sobre os mesmos.

4.4.2.4. Conclusões

A partir dos resultados apresentados conclui-se: o perfil do público que avaliou o Método encontra-se dentro do tipo de público previsto para o perfil do usuário.

Os resultados da avaliação realizada na aplicação do modelo TAM, durante a reunião do Grupo Focal, são convergentes com as observações adicionais realizadas pelo mesmo Grupo Focal durante o segundo momento da dinâmica. O Método se mostra útil e aplicável, porém a facilidade de uso não é satisfatória, sem chegar a ser ruim.

A complexidade do Método é sem dúvida ponto mais crítico e impacta diretamente na facilidade de uso e se manifesta nos seguintes aspectos: tempo de duração do Tutorial e o fato do Tutorial ser cansativo de demandar um grande esforço do usuário. Cabe destacar a Facilidade de Uso certamente viu-se impactada pelas condições em que foi feita a avaliação. Em uma aplicação rotineira dificilmente o

usuário usaria o Tutorial em forma ininterrupta durante uma hora como de fato aconteceu durante a avaliação. Ainda a avaliação foi realizada durante a noite, após os participantes terem executado sua jornada regular de trabalho.

A aplicabilidade e utilidade do Método é afirmada por comentários realizados pelos avaliadores como o ensejo de usar o Tutorial como forma de treinamento das equipes das suas organizações.

4.5. AÇÕES

A avaliação inicial dos especialistas gerou algumas ações de melhoria para o Tutorial e para os questionários que foram usados na avaliação realizada pelo Grupo Focal, estas ações estão descritas na Subseção 4.4.1. Em relação à avaliação realizada pelo Grupo Focal a complexidade do Método foi identificada como a questão mais crítica e que impacta diretamente em diversos aspectos vinculados à facilidade de uso.

A mineração de dados é um processo de um alto grau de complexidade, este está inserido no processo de KDD que faz parte do Método. Ainda como parte do Método deve-se considerar a análise do contexto de aplicação, necessária para garantir o cumprimento do requisito do Método de preservar a operação do BI. Também se tem como requisito do Método é que este execute um conjunto de ações que levem à transição do estado inicial para o final passo a passo. Baixo estas considerações o Método acaba incorporando a complexidade do processo de Mineração, acrescido das atividades inerentes a KDD e das atividades de validação do contexto da sua aplicação. Assim sendo o Método se torna longo e complexo. Na concepção do Método questões como geração de relatórios que acompanhem a execução e apoio dos experts foram pensadas como forma de mitigar estas dificuldades

Perante a constatação da complexidade do Método, algumas melhorias podem ser realizadas no mesmo. Estas melhorias estão vinculadas à sua estrutura, organização das atividades ou mesmo à forma como o Método é apresentado. Porém, estas mudanças não estão sendo consideradas dentro do escopo deste trabalho por limitações de tempo para a conclusão do mesmo.

5. EXERCÍCIO DE APLICAÇÃO DO MÉTODO.

Neste Capítulo apresenta-se um exercício de aplicação do Método. Para execução do exercício conta-se com um modelo OLAP. O Modelo OLAP representa uma visão organizada de grandes volumes de dados. Em particular o modelo OLAP usada no exercício apresenta o formato estrela.

Os dados que compõem o modelo são dados abertos disponíveis no *site* da Prefeitura de Porto Alegre. Estes dados contêm informações referentes a despesas orçamentárias geradas no período 2010-2014. A Prefeitura de Porto Alegre disponibiliza ao cidadão um conjunto de consultas baseadas em dados históricos com o objetivo oferecer transparência sobre as despesas realizadas.

As despesas estão bem exploradas pelo ponto de vista de análise do passado. Contudo, nada se tem para prever o futuro, principalmente para auxiliar no planejamento das despesas com antecedência e com boa margem de acerto. Também deseja-se conhecer a composição orçamentária de acordo com os principais elementos que definem o orçamento (Unidade Orçamentária, Processo de Contratação, Tipo de Empenho, etc.). Para atingir este objetivo, será necessário aplicar mecanismo de Descoberta de Conhecimento em Banco de Dados, em particular gerar Modelos de Previsão.

Para execução deste exercício foi usado o Tutorial, Apêndice A, seguindo suas instruções. Cada uma das ações executadas foi registrada em uma planilha indicando a ação que foi executada e qual foi o resultado obtido. A definição completa do exercício assim como a sua execução, restrições para sua execução e resultados, encontram-se no Apêndice B.

5.1. CONSIDERAÇÕES SOBRE A APLICAÇÃO DO MÉTODO.

O exercício de aplicação prática do Método apresentou algumas restrições para sua execução:

- Não foi realizado sobre uma plataforma DW/BI, fato que impossibilitou a execução das ações referentes ao conhecimento do contexto de aplicação e avaliação da análise de impacto.

- A aplicação foi realizada pela autora em um laboratório fechado o que impossibilitou a interação com os diferentes especialistas direcionados pelo Método.
- A não presença do Negócio criou barreiras para a autora na execução de ações como: conhecimento dos dados, identificação dos critérios de sucesso do Modelo, interpretação dos padrões e outras atividades, onde o papel do Negócio é muito importante.
- O Modelo OLAP obtido para o exercício passou por um processo de tratamento de dados não conhecido.
- O uso de ferramentas livres, de código aberto para a geração do Modelo de Dados, dificultou o trabalho devido à falta de documentação das suas funcionalidades.

A aplicação exercício deu lugar a um conjunto de considerações:

Sobre o Método em geral

- O Método efetivamente direciona, passo a passo, para o resultado desejado.
- O uso do Tutorial é de alta relevância para acompanhar a execução do Método dada a sua complexidade.
- A aplicação do Método não é simples e envolve um grande volume de informações.
- O apoio dos especialistas é necessário dada a tangência do Método com diversos domínios de conhecimento que não fazem parte dos domínios de conhecimento definidos para o Usuário do Método.
- A interação com o Negócio é de extrema importância para otimizar e inclusive viabilizar a execução do Método. O conhecimento do Negócio impacta diretamente na definição dos aspectos computacionais que fazem parte do Método.
- A criação de relatórios para registrar informações relevantes e acompanhar a execução do Método é necessária dado o grande volume de informações geradas.
- A identificação e o estudo comparativo de procedimentos semelhantes que são aplicados no estado inicial e ao mesmo tempo no estado final, se torna imprescindível para poder garantir os objetivos buscados na execução do

Método. Isto se deve a que os procedimentos realizados no estado inicial geram premissas para os procedimentos semelhantes que serão executados no estado final .

Sobre aspectos referentes a detalhes da execução:

- O relatório de Especificações do Método precisa ser reestruturado para implementar a geração de mais de um Modelo de Dados.
- O Método deve contemplar a geração de visualizações como forma de implementar algumas tarefas. Em tal caso deve existir registro dos parâmetros usados para gerar as visualizações.
- Deve-se registrar a parametrização usada na ferramenta para geração dos diferentes Modelos. Diferentes parâmetros geram Modelos diferentes e podem até gerar novas restrições de HW.
- Deve-se estudar a possibilidade de unir as informações sobre os Dados com as informações sobre o Modelo de Dados. Na medida que se avança no Método estes dois grupos de informações ficam cada vez mais dependentes.
- Deve ser desenhada uma tabela padrão para incluir os dados dos atributos, informações sobre como foram tratados, etc. Esta tabela deve fazer parte do Relatório de Dados.
- A avaliação constante da infraestrutura de HW para suporte à formação do Modelo deve ser colocada em forma explicita como parte das ações que implementam o Modelo. O HW é impactado pelas diferentes parametrizações que podem ser usadas na ferramenta e que, por sua vez, também se veem impactadas pelos dados que formam o Modelo. Estes impactos apenas são constatados em tempo de formação do Modelo de Dados. Exemplo: uso da memória na geração das árvores. Diferentes opções de poda podem ser usadas.
- No uso de ferramentas livres e de código aberto deve-se dar especial atenção às versões usadas e à documentação disponível para a mesma. Versões mais recentes carecem de documentação ou até mesmo de registros em fóruns onde este conhecimento é compartilhado.

- Para a formação do Modelo deve-se conhecer o tratamento de dados já realizado no ETL e na formação do modelo OLAP. Estas informações devem fazer parte das informações sobre os dados.
- A Tarefa 6 tem como objetivo realizar uma avaliação geral da execução do Método. Para execução desta tarefa o Método indica a interação com o Negócio. Porém, a interação com todos os envolvidos certamente seria muito mais proveitosa, já que cada um poderia dar um parecer sobre a sua participação.

Mudanças no Método visando a incorporação das considerações acima assinaladas, não é fazer parte do escopo deste trabalho.

6. TRABALHOS RELACIONADOS

O objetivo deste Capítulo é apresentar artigos acadêmicos que possuem similaridades com o trabalho que está sendo proposto.

Robert [ROB11] coloca que conceitualmente o termo “maturidade” descreve um “estado de ser completo, perfeito ou pronto”. Para alcançar um estado de maturidade desejado, um caminho evolutivo de um estágio inicial para um estágio alvo deve ser seguido. Cada estágio é definido pelo nível de desenvolvimento de um conjunto de elementos organizacionais chamados de dimensões. Diferentes modelos podem ser usados para orientar esse processo de transformação. Estes modelos são chamados de Modelos de Maturidade. Continuando, Robert [ROB11] apresenta uma avaliação de diferentes Modelos de Maturidade para BI fazendo inicialmente algumas considerações gerais. Deve-se destacar a afirmativa realizada pelo autor “Os Modelos de Maturidade revelam o caráter dinâmico da Inteligência Organizacional, sendo este um processo em constante melhoria, a maturidade é um sinal de superior desempenho organizacional”.

A Universidade de Carnegie Mellon [CMM14] apresenta o *Data Maturity Model* – DMM. O DMM é definido como um conjunto de processos que implementam as funções de construção, manutenção e conformidade para os ativos de dados. O objetivo principal é gerar a estrutura administrativa e operacional necessária para garantir que os dados corporativos sejam gerenciados como um ativo crítico para o desenvolvimento organizacional. O modelo está composto por cinco estágios, no qual o primeiro descreve os dados como “requisitos para implementação dos projetos”, evoluindo até o quinto estágio onde os dados vêm-se como “críticos para a sobrevivência da organização e competitividade no mercado”.

Mendoça Xavier et al. [XAV16] apresenta uma análise comparativa de diversos Modelos de Maturidade para dados.

Cguan [CGU10] baseia-se na análise de modelos já existentes para BI e propõe um novo modelo para o referido contexto o *Enterprise Business Intelligence Maturity Model* - EBIMM. O autor identifica em cada nível evolutivo um conjunto de elementos que considera críticos e ao mesmo tempo identificam o nível evolutivo. Na medida que se evolui nos diferentes níveis de maturidade a junção dos elementos críticos se torna mais complexa.

Lahrmann et al.[LAH11] desenvolveu um modelo teórico para organizações que usam BI. Este modelo integra aspectos tecnológicos e de negócio para avaliar o nível de maturidade no uso da referida tecnologia. O nível de maturidade é determinado pela forma como o BI está implementado e o impacto este que gera no negócio. De acordo com o autor, este modelo pode ser considerado um guia para implementar BI dentro da organização, de forma que esta tecnologia contribua para a obtenção de melhores resultados na organização.

Chen et al. [CHE16], em seu estudo, investiga como uma organização precisa se preparar para poder usar com sucesso tecnologia de BI e identificar as melhores práticas a aplicar para colher os melhores resultados.

Concluindo, pode-se afirmar que existem diversos trabalhos publicados que apresentam Modelos de Maturidade que tratam da transição das organizações de um estado atual para outro desejado, na procura de um melhor aproveitamento dos dados em prol de um comportamento organizacional diferenciado. Em particular, alguns dos trabalhos apresentados consideram organizações usuárias de BI. Destaca-se a crescente complexidade presente na evolução de um nível a outro. Esta complexidade é dada por um conjunto de elementos envolvidos na definição de cada estado de maturidade que elenca aspectos computacionais e do negócio. Também pode-se constatar que existe um ponto de partida ou de prontidão, no qual coexistem um conjunto de premissas que a organização precisa satisfazer para poder iniciar o desenvolvimento da sua maturidade no uso de dados. Porém, nenhum destes trabalhos trata especificamente em como fazer a transição de um estado onde os dados são usados para a tomada de decisão para outro novo estado onde são aplicadas outras formas mais complexas de análise de dados. Ainda esta transição deve focar em questões computacionais e manter operacional a base BI. Esta mudança permitirá um melhor aproveitamento dos dados para obtenção de um comportamento organizacional diferenciado de forma que caracterize à organização onde é aplicado como uma Organização com Comportamento Inteligente.

7. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O Método proposto neste trabalho implementa uma sequência organizada de atividades que possibilita que organizações usuárias de BI possam fazer a transição para a aplicação de KDD sem que isto afete a operação da solução BI já instalada. Partindo de requisitos fornecidos pelo Negócio e mantendo o foco em questões computacionais, o Método entrega ao Negócio novos padrões de dados para sua avaliação e posterior aplicação. As questões computacionais contempladas envolvem: identificação de requisitos, seleção e tratamento de dados, construção do Modelo de Dados, seleção e aplicação de ferramentas para a Construção do Modelo de Dados, entrega de resultados. O Modelo mantém a restrição inicial de mencionar aspectos do negócio apenas como forma de contextualizar os aspectos computacionais, também no papel de geradores de requisitos, avaliadores dos produtos gerados e fornecedores de conhecimento de apoio a todo o processo computacional.

Para avaliar o Método foi criado um tutorial e este submetido a um conjunto de experts que integraram um Grupo Focal. Como ferramenta de avaliação o Grupo Focal respondeu a um questionário gerado de acordo com o modelo TAM. O Grupo Focal também contribuiu com uma série de observações. A avaliação se mostrou convergente em seu resultado apontado o Método como aplicável, porém sua facilidade de uso não se mostrou satisfatória mostrando-se como altamente complexo.

A complexidade do Método resulta do próprio processo de KDD. Este processo além de ser interativo e iterativo, reúne um alto número de variáveis que se elencam de diferentes formas na medida que o processo evolui na sua execução. O fato de o Método estar sendo aplicado em uma plataforma DW/BI, cria um conjunto de premissas e restrições para sua execução que o tornam ainda mais complexo. A implementação do Método em ações que viabilizam sua execução passo a passo permite que o objetivo estabelecido seja atingido. Porém, esta característica torna o Método lento e cansativo na sua aplicação.

O exercício de aplicação do Método mostrou que o mesmo é eficiente na sua proposta de estar estruturado por uma sequência organizada de atividades que concretizam a sua implementação, cobrindo os aspectos computacionais enunciados na sua definição. O exercício também mostrou que o Modelo oferece alguns atenuantes para mitigar a sua complexidade. O uso de relatórios que condensam

informações críticas e acompanham a execução das atividades se mostrou como uma ferramenta muito valiosa de apoio. Outro atenuante que o Método apresenta é a indicação de especialistas que podem apoiar o usuário durante a execução. A tangência das atividades do Método com outros domínios de conhecimento, que não fazem parte dos domínios de conhecimento do usuário do Método, dão grande relevância à participação dos especialistas.

A interação com o Negócio mostrou-se com um ponto fundamental para a execução do Método. Mesmo que o escopo do trabalho está restrito a questões computacionais, atividades que envolvem o Negócio iniciam e finaliza a execução do Método. O Negócio também é foco de inúmeras iterações durante a execução, gerando impactos na definição dos diferentes aspectos computacionais abordados.

Foi constatado que um ponto extremamente importante para a eficiente execução do Método é conhecer os procedimentos semelhantes que são realizados na plataforma DW/BI e no processo de KDD. Os procedimentos realizados na plataforma DW/BI geram premissas para os procedimentos realizados no processo de KDD e que constituem o Método. O não conhecimento destas premissas pode comprometer os resultados buscados na aplicação do Método.

A partir das colocações acima realizadas, surgem algumas demandas para trabalhos futuros:

- Devido ao resultado da avaliação referente à facilidade de uso que mostra o Método como complexo, deve-se pensar em novos mecanismos que ajudem a reduzir ou mitigar a complexidade do Método.
- Devido à relevância que o Negócio apresenta na aplicação do Método e ao forte vínculo do Negócio com as questões computacionais, considera-se em ampliar o escopo de definição do Método para incluir uma abordagem profunda dos aspectos do Negócio.
- Devido ao impacto gerado na construção do Modelo de Dados, considera-se importante para desenvolvimento é incluir como parte do Método uma análise detalhada e comparativa dos procedimentos de tratamento de dados executados para a formação do Modelo de Dados BI e no Modelo de Dados de KDD.

REFERÊNCIAS

- [ANG03] ANGELONI, M.T. “Elementos intervenientes na tomada de decisão”. Revista SCIELO, Janeiro 2003, pp. 17-22.
- [ASC07] ASCACIBAR F.M. “*Optimización Mediante Técnicas de Mineración de Datos Del Ciclo Reconocido de una Linea de Galvanizado*”, Tese de Doutorado, Universidad de La Rioja, Janeiro 2007, 644p.
- [BAR09] BARBOUR R. “Grupos Focais”. Porto Alegre: Artmed, 2009, 206p.
- [BRE04] BRESLIN M. “*Comparing the basics of Kimball and Inmon*”. *Business Intelligence Journal*, Inverno 2004, pp. 6-20.
- [CAO17] CAO L. “*Data Science: A Comprehensive Overview*”. *Communications of the ACM*, Agosto 2017, pp 59-68.
- [CAS12] CASALI A.; ERNST C. “*Discovering Correlated Parameters in Semiconductor Manufacturing Processes: A Data Mining Approach*”. *IEEE Transactions on Semiconductor Manufacturing*, Fevereiro 2012, pp.118-127.
- [CGU10] CGUAN M. H. “*An enterprise business intelligence maturity model (EBIMM): Conceptual Framework* “. In: *Proceedings of the Fifth International Conference on Digital Information Management*, 2010, pp. 303-308.
- [CHA00] CHAPMAN P.; CLINTON J.; KERBER R.; KHABAZA T.; REINARTZ T.; SHEARER C.; WIRTH R. “*CRISP-DM 1.0 - Step-by-step data mining guide*”. Recuperado de: <http://www.crisp-dm.org/CRISPWP-0800> , Novembro 2017.
- [CHA98] CHAN P. K.; STOLFO S.J. “*Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection*”. In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, 1998, pp 164-168.
- [CHE16] CHEN X.; CHEN L.; CLAIRE E.; BAJAWA D. “*Organizational Readiness for Business Intelligence and Analytics Systems Success Emergent Research*”, In: *Proceedings of the Twenty-second Americas Conference on Information Systems*, 2016, pp 1-5.
- [CHO06] CHOO C.W. “*A Organização do Conhecimento*”. São Paulo: Senac,2006, 421p.
- [CHU10] CGUAN M. “*An enterprise business intelligence maturity model (EBIMM): Conceptual Framework* “. In: *Proceedings of the Fifth International Conference on Digital Information Management*, 2010, pp. 303-308.
- [CMM14] CARNIEGIE MELLON UNIVERSITY. 2017. “*CMMI - Capability Maturity Model, Capability Maturity Modeling*”. Recuperado de: www.sei.cmu.edu/legal/marks/index.cfm, Novembro 2017, 53p.
- [DAV89] DAVIS F. “*Perceived Usefulness, Perceiver Ease of Use, and User Acceptance of Information Technology*”. *MIS Quarterly*, Setembro 1989, pp. 319-340.

- [FAY96] FAYYAD U.; SHAPIRO G.P.; AMYTH P. "*From Data Mining to Knowledge Discovery in Databases*". AI Magazine, Março 1996, pp. 37-54.
- [FRE13] FEITAS C. A.; SCONS F.E.; "Um Sistema de Informação para apoio à elaboração e realização de provas no Projeto Viva voz". Trabalho de Conclusão, PUCRS, Dezembro 2013, 121p.
- [GAG15] GAGNON R.; "*Betting on Big Data how the right culture. Strategy and investments can help you leapfrog the competition*". Forbes Insight, Setembro 2015, pp. 1-30.
- [GAR98] GARDNER H.; HORNHABER M. J.; WAKE W.K. "Inteligência – Múltiplas perspectivas". Porto Alegre: Artmed, 1998, 356p.
- [HAA14] HAAS L. "*The IBM Research Accelerated Discovery Lab*". Newsletter ACM ,Dezembro 2014, pp. 41-48.
- [HAM14] HAMMAD A.; ABOURZIK S.; MOHAMED Y. "*Application of KDD Techniques to Extract Useful Knowledge from Labor Resources Data in Industrial Construction Projects*". Journal of Management in Engineering, Novembro 2014, pp. 1-10.
- [HIN07] HINOJOSA L. M.; J.A. P. MORENO J. A. "*Manual práctico para el diseño de la Escala Likert*". San Agustín Tlaxiaca: Trillas, 2007, 97p.
- [KIM13] KIMBALL R.; ROSS M; "*The Data Warehouse Toolkit*". Inidianópolis: John Wiley & Sons Inc., 2013, 563p.
- [KIR12] KIRON D.; FERGUSON R.B.; PRENTICE P.K. "*Innovating with Analytics*". MITSloan Management Review, Setembro 2012, pp. 47-52.
- [KIR13] KIRON D.; FERGUSON R. B.; PRENTICE P.K. "*From value to Vision, reimaginig the possible with Data Analytics*". MITSloan Management Review, Research Report, Março 2013, 22p.
- [LAH11] LAHRMANN G.; F. MARX F.; WINTER R.; WORTMANN F. "*Business Intelligence Maturity: Development and Evaluation of a Theoretical Model*". In: *Proceedings of the 44th Hawaii International Conference on System Sciences*, 2011, pp. 206-246.
- [LEV00] LEVY P. "A inteligência Coletiva". São Paulo: Edições Layala, 2000, 210p.
- [MAE06] MAES A.; POELS G. "*Evaluating Quality of Conceptual Models Based on User Perceptions*". In: *Proceedings of the International Conference on Conceptual Modeling*, 2006, pp. 54-67.

- [MAR10] MARISCAL G.; MARABÁN O.; FERNANDEZ C. “*A survey of data mining and knowledge discovery process models and methodologies*”. The Knowledge Engineering Review, Junho 2010, pp. 137–166.
- [MCM16] Mc MASTER M. “*What is Organizational Intelligence?*”. USA: Space Independent Publishing Platform, 2016, 103p.
- [MOR96] MORGAN D. L., “*Focus Groups*”. Annual Review of Sociology, Agosto 1996, pp. 129-152.
- [MUL12] MULLER R.; CASTILLO N. C. “*Inteligência Organizacional como ferramenta de Gestão: um referencial teórico integrado*”. Revista Expectativa, Janeiro 2012, pp. 83-102.
- [NIA15] NIAKSU O. “*CRISP Data Mining Methodology Extension for Medical Domain*”. Baltic J. Modern Computing, Janeiro 2015, pp. 92-109.
- [NON08] NONAKA I.; TAKEUCHI H. “*Gestão do Conhecimento*”. Porto Alegre: Bookman, 2008, 320p.
- [PET11] PETERSON F. C. “*Processo de indução e ranqueamento de árvores de decisão sobre modelos OLTP*”. Dissertação de Mestrado, PUCRS, Março 2011, 109p.
- [PLU17] PLUMES F.M.; “*CASP-DM: Context Aware Standard Process for Data Mining*”. Cornell University Library, Setembro 2017, 38p.
- [RIV15] RIVEIRO L.; VICENZI A.; MALDONADO J. C.; T. CONTE T. “*Evaluating Software Engineers’ Acceptance of a Technique and Tool for Web Usability Inspection*”. In: *Proceedings of the 7th International Conference on Software Engineering and Knowledge Engineering*, 2015, pp. 114 – 121.
- [RAH14] RAHMAN F. A. “*Data Cleaning in Knowledge Discovery Database (KDD)-Data Mining*”. In: *Proceedings of the International Conference on Advance in Computer Science and Electronics Engineering*, 2014, pp 4-10.
- [RAN18] RANSBOTHAM S.; KIRON D. “*Findings from the 2018 data & Analytics Global Executive Study and Research Project*”. MITSMR Report, Winter 2018, 22p.
- [ROB11] ROBERT F. M. “*Business Intelligence Maturity: Development and Evaluation of a Theoretical Model*”. In: *Proceedings of the 44th Hawaii International Conference on System Sciences*, 2011, pp 1-10.
- [RUS11] RUSSOM P. “*Data Analytics Fourth Quarter 2011*”. TDWI Best Practice Report, Fouth Quarter 2011, 36p.

- [RUS13] RUSSOM P. “Gerenciando Dados Principais, Relatório de Melhores Práticas do TDWI, Q4”. Recuperado de: <http://www.tdwi.org/bpreports>, Novembro 2017.
- [SAB07] SABBAG P. Y. “Espirais do Conhecimento”. São Paulo: Saraiva, 2007, 350p.
- [SCH03] SCHVARSTEI L. “*La inteligencia social de las organizaciones - desarrollando las competencias necesarias para el ejercicio efectivo de la responsabilidad social*”. Argentina: Paidós, 2003, 272p.
- [SHA08] SHRAMA S.; OSEI-BRYSON K. “*Organization-Ontology Based Framework for Implementing the Business Understanding Phase of Data Mining Projects*”. In: *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, 2008, pp. 1-77.
- [VIR14] VIRGOLIM A. M. R., “Altas habilidades/superdotação, inteligência e criatividade: uma visão multidisciplinar”. Campinas: Papyrus, 2014, 224p.
- [WIL15] WILENSKY H. L. “*Organizational Intelligence: Knowledge and Policy in Government and Industry (Classics of the Social Sciences)*”. USA: QP Books, 2015, 6043p.
- [XAV16] MENDONÇA XAVIER J. E.; R. A. MARTINS R. A. “Análise dos Modelos de Maturidade Analítica”. In: *Proceedings of the XXXVI Encontro Nacional de Engenharia de Produção*, 2016, 11p.

APÊNDICE A

O Tutorial

SLIDE 1

Este *slide* tem como objetivo organizar a apresentação.



SLIDE 2

Como funciona o tutorial?

O Tutorial está composto por um conjunto de ações que implementam um Método. Este Método permite ao Analista de BI **executar o processo de Descoberta de Conhecimento em Banco de Dados** sobre a plataforma DW/BI, conservando as funcionalidades da mesma.

Na sua concepção, o Método foi pensado para sua aplicação em diferentes áreas do Negócio, desde que a área de negócio use Sistemas de Apoio à Tomada de Decisão gerados a partir de dados que fazem parte de uma plataforma que integra *Business Intelligence* e *Data Warehouse*. O ponto de partida e de chegada para a execução do Método é o Negócio. O entendimento da necessidade do Negócio em

relação à obtenção do novo conhecimento direciona todas as atividades que compõem o Método. Finalizando, o Negócio avalia e valida os resultados obtidos.

As Tarefas que compõem o Método se desdobram em Atividades, que por sua vez estão compostas por um conjunto de ações executáveis. Muitas destas Atividades apresentam informações adicionais para facilitar seu entendimento e implementação. Pelo fato de o Método estar sendo aplicado sobre uma plataforma DW/BI, seus componentes e funcionalidades são continuamente referenciados em função dos impactos que a aplicação do Método pode gerar na plataforma.

Cada *slide* é acompanhado de um relato, o relato inicia clicando no ícone que representa o som e pode ser escutado reiteradamente antes de passar para o próximo *slide*. A passagem de um *slide* para outro é realizada pelo usuário, permitindo que ele evolua na apresentação de acordo com seu próprio ritmo.

O público alvo para este Tutorial é o Analista de *BI*. Considera-se que Analista de BI tem como responsabilidades:

- Garantir que o ambiente de BI atenda tecnicamente às necessidades do negócio.
- Desenvolver a fazer a manutenção das aplicações de BI.
- Configurar das ferramentas de BI.
- Manejar o banco de dados.
- Entender o negócio e de seus dados subjacentes.
- Desenvolver a arquitetura de dados garantindo sua usabilidade, integração e otimização.
- Executar a análise detalhada dos dados para desenvolvimento do Modelo de dados de BI.

Na tabela com o título “Conhecimento”, apresentada no *slide*, identificam-se os domínios de conhecimento e nível de profundidade referente aos mesmos que o Analista de *BI* deve ter para poder aplicar o Método descrito neste Tutorial. Em alguns domínios de conhecimento, o Analista de BI poderá precisar do apoio de especialistas. Em outro *slide* apresenta-se um glossário contendo as siglas usadas no Tutorial. Detenha-se na leitura do glossário para garantir o entendimento do Tutorial.

1 

COMO FUNCIONA O TUTORIAL?



Objetivo

O Tutorial apresenta um conjunto de Tarefas que compõem um **Método** que permite ao Analista de BI **executar o processo de Descoberta de Conhecimento em Banco de dados** sobre a estrutura DW/BI, conservando as funcionalidades da mesma.



Público-alvo
Analistas de BI



Conhecimento

Domínio de Conhecimento	Nível
Técnica para interação com o Negócio	BÁSICO
Administração de Dados no <i>Data Warehouse</i>	BÁSICO
Arquitetura e desenvolvimento do ETL	BÁSICO
Arquitetura e infraestrutura da plataforma DW/BI	BÁSICO
Administração dos Metadados na plataforma DW/BI	AVANÇADO
Técnicas e Ferramentas de Mineração de Dados	BÁSICO
Desenvolvimento de Aplicações BI	AVANÇADO
Ferramentas de BI	AVANÇADO
Arquitetura de Dados	AVANÇADO
Preparação e Modelagem de Dados para BI	AVANÇADO

SLIDE 3

Como funciona o tutorial?

O Tutorial está ilustrado por ícones que sinalizam informações complementares, não por isso menos importantes, referentes à execução do Método. De acordo com a sequência apresentada no *slide*, descreve-se o significado de cada ícone:

Interação: o Método é totalmente interativo, demandando o apoio de outros especialistas em domínios de conhecimento que podem não fazer parte do conhecimento do Analista de BI. Estes domínios de conhecimento dentro da organização se traduzem em papéis. Estes papéis podem estar representados pela mesma pessoa ou não, variando de organização para organização. O Tutorial se propõe oferecer as informações necessárias para que o Analista de BI, usuário deste Método, possa interagir com estes especialistas quando necessário. Lembrando que em alguns casos o próprio Analista de BI pode ser ao mesmo tempo especialista em outros domínios de conhecimento. Em outro *slide*, a continuação, são apresentados estes especialistas e suas responsabilidades.

Iteração: o Método é iterativo. Isto significa que de acordo com o resultado da execução de algumas atividades, pode ser necessário retornar a tarefas anteriores onde é reiniciada a execução do Método. O ícone que representa a iteração é acompanhado da condição que precisa ser avaliada e da atividade onde o Método deve reiniciar, caso seja necessário.

Saídas: em todas as tarefas existe uma indicação do conjunto de saídas que elas produzem como resultado da execução de suas atividades. As saídas indicadas pelo Método **devem** ser geradas. Isto é necessário para documentar e posteriormente facilitar a execução das atividades. Ainda, enfatizando a importância de efetivamente gerar as saídas, cabe destacar que as informações documentadas se transformam em entradas para outras tarefas ou atividades do Método.

Entradas: acompanhando a definição de cada tarefa apresentam-se as entradas para suas atividades. Posteriormente, estas entradas estão particularmente identificadas para cada atividade. Não deixe de considerar estas entradas durante a execução das atividades, as informações ou produtos apresentados na mesma são fundamentais para garantir a correta execução das ações e assim facilitar a aplicação do Método.

Som: clicando neste ícone você poderá escutar a explicação que acompanha alguns dos *slides* que compõem este Tutorial. Você poderá interagir com o Tutorial e escutar o relato associado a um *slide* quantas vezes considerar necessário.

Alguns *slides* não possuem informações já que eles se propõem apenas a organizar as informações. Existem outro tipo de *slides* que não estão acompanhados de scripts porque consideram-se totalmente autoexplicativos. Estes dois tipos de *slides* não contêm o ícone de som.



SLIDE 4

Com quem interagir?

Os papéis apresentados representam os responsáveis por procedimentos executados no desenvolvimento das atividades associadas ao contexto da plataforma DW/BI. A distribuição destes papéis em diferentes pessoas varia de uma organização para outra.

Existem outros papéis no contexto da plataforma *DW/BI* além dos apresentados. Porém, neste trabalho apenas são citados aqueles papéis que de uma forma ou outra têm participação direta na aplicação do Método.

Analista de Negócios: o Analista de Negócios é responsável por liderar o processo de definição dos requisitos do Negócio e trabalhar como representante destas informações frente à equipe técnica.

Administrador de Dados: o Administrador de Dados é responsável por direcionar a organização na definição de regras do Negócio e domínio de valores permitidos para os dados que compõem o *Data Warehouse*. O Administrador de Dados também é responsável pela qualidade de dados carregados no *Data Warehouse*.

Arquiteto de ETL: o Arquiteto de ETL é responsável pelo processo de extrair, transformar e carregar os dados no *Data Warehouse*. O Arquiteto de ETL precisa conhecer os sistemas onde os dados se originam e entender o modelo de dados do qual farão parte.

Arquiteto Técnico: o Arquiteto Técnico é responsável pelo desenho da arquitetura técnica e o suporte técnico para a plataforma *DW/BI*. Deve garantir que todos os elementos que compõem a arquitetura possam coexistir adequadamente. Tem um papel fundamental na avaliação e seleção dos produtos componentes assim como na sua instalação.

Gestor de Metadados: o Gestor de Metadados é responsável por todos os metadados que fazem parte da plataforma *DW/BI*, desde sua identificação e cadastro até a forma como são apresentados para a comunidade do negócio.

Especialista em Mineração: o Especialista em Mineração é responsável por desenvolver os modelos de dados aplicando metodologias e ferramentas de mineração de dados. Deve ter conhecimento de estatística e do negócio.



COM
QUEM
INTERAGIR?

3 

ANALISTA DE NEGÓCIOS
Lídera o processo de definição de requisitos

ADMINISTRADOR DE DADOS
Gestor dos dados do Data Warehouse

ARQUITETO DE ETL
Arquitetura/desenvolvimento do ETL

ARQUITETO TÉCNICO
Arquitetura e Infraestrutura da plataforma DW/BI

GESTOR DE METADADOS
Administração dos Metadados na plataforma DW/BI

ESPECIALISTA EM MINERAÇÃO
Técnicas e ferramentas de Mineração

SLIDE 5

Este *slide* tem como objetivo apresentar o glossário.

GLOSSÁRIO

BI – *Business Intelligence*

DW/BI - *Data Warehouse/Business Intelligence*

DW – *Data Warehouse*

ETL - *Extract, Transform and Load*

KDD – *Knowledge Discovery in Databases*

OLAP - *On-line Analytical Processing*

SAD – *Sistema de Apoio à Decisão* SQL - *Structured Query Language*

SLIDE 6

O QUE É KDD?
(Knowledge Discovery in Database – Descoberta de Conhecimento em Banco de Dados)

“

KDD é um processo que essencialmente garante que **conhecimento útil** seja extraído de dados. A capacidade das organizações em gerar, processar e aplicar este conhecimento determina o **crescimento da sua “força mental”**. Esta “força mental” permite à organização manifestar um **comportamento diferenciado** em prol da sua missão. Este comportamento caracteriza as chamadas **Organizações Inteligentes.**

”

(FAYYAD, 1996)

SLIDE 7

KDD (Knowledge Discovery in Databases)

O processo de extração de conhecimento a partir de grandes conjuntos de dados é um processo não trivial, interativo e iterativo. A complexidade deste processo está principalmente na dificuldade de interpretar os resultados da aplicação de cada uma das suas tarefas e conjugar estas interpretações da forma em que devem ser executadas as tarefas definidas na continuação.

O processo de KDD é rotulado como interativo, indicando a necessidade da atuação eventual de um especialista como responsável do controle do processo. O especialista aplica os recursos computacionais e do Negócio em função dos fatos observados e dos resultados envolvidos. O termo iterativo indica a possibilidade de repetições das atividades do processo em busca de resultados satisfatórios.

A aplicação do processo de KDD permite identificar Padrões. Estes devem ser interpretados pelo homem para se converterem em conhecimento. Um Padrão descreve fatos ou tendências associadas a um conjunto de dados. Os Padrões podem ser classificados em 2 tipos básicos: preditivos e descritivos. Os Padrões Preditivos são construídos com o intuito de resolver o problema específico de prever valores de um ou mais atributos. Os Padrões Descritivos apresentam informações que um especialista do domínio da aplicação pode não conhecer. Os padrões encontrados como o resultado da aplicação do processo de KDD devem atender a 4 critérios. Eles devem ser:

- Compreensíveis: permitem um entendimento claro pelos usuários do domínio da aplicação, isto implica uma representação simples que os torne inteligíveis. Uma técnica possível para atingir este objetivo é a visualização.
- Válidos: o conhecimento deve ser adequado ao contexto de aplicação.
- Novos: devem acrescentar novos conhecimentos aos já existentes no contexto de aplicação e principalmente para o especialista do domínio.
- Úteis: podem ser aplicados de forma a proporcionar benefícios ao contexto da aplicação do KDD.



SLIDE 8

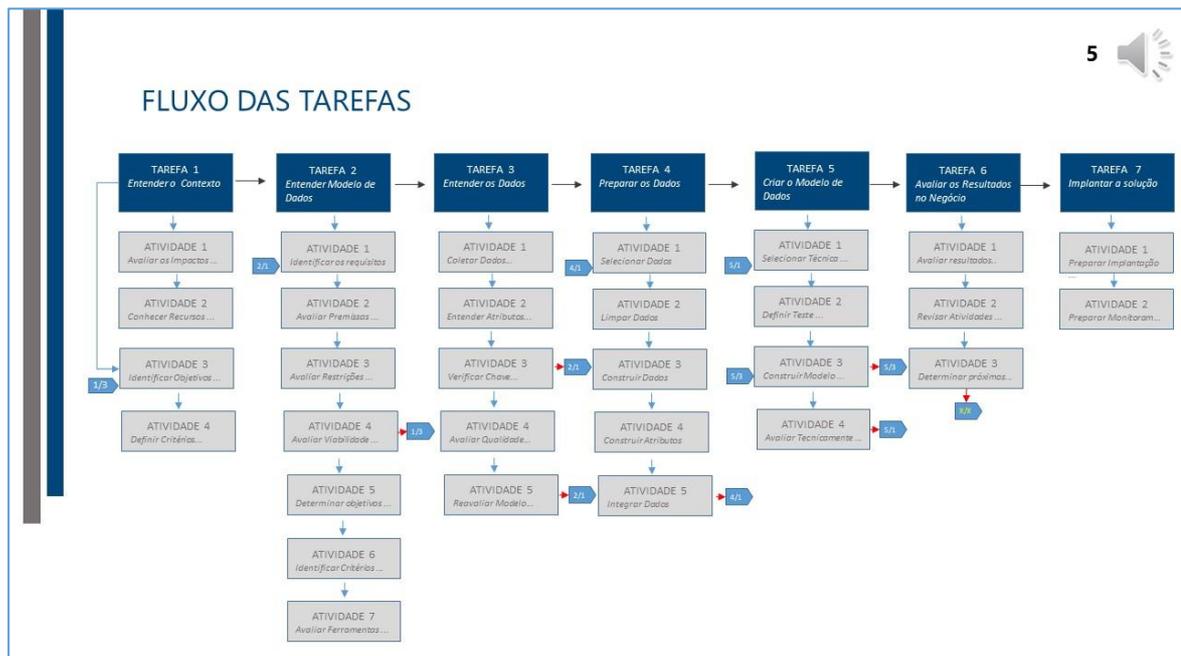
Este *slide* tem como objetivo organizar a apresentação.

TAREFAS

SLIDE 9

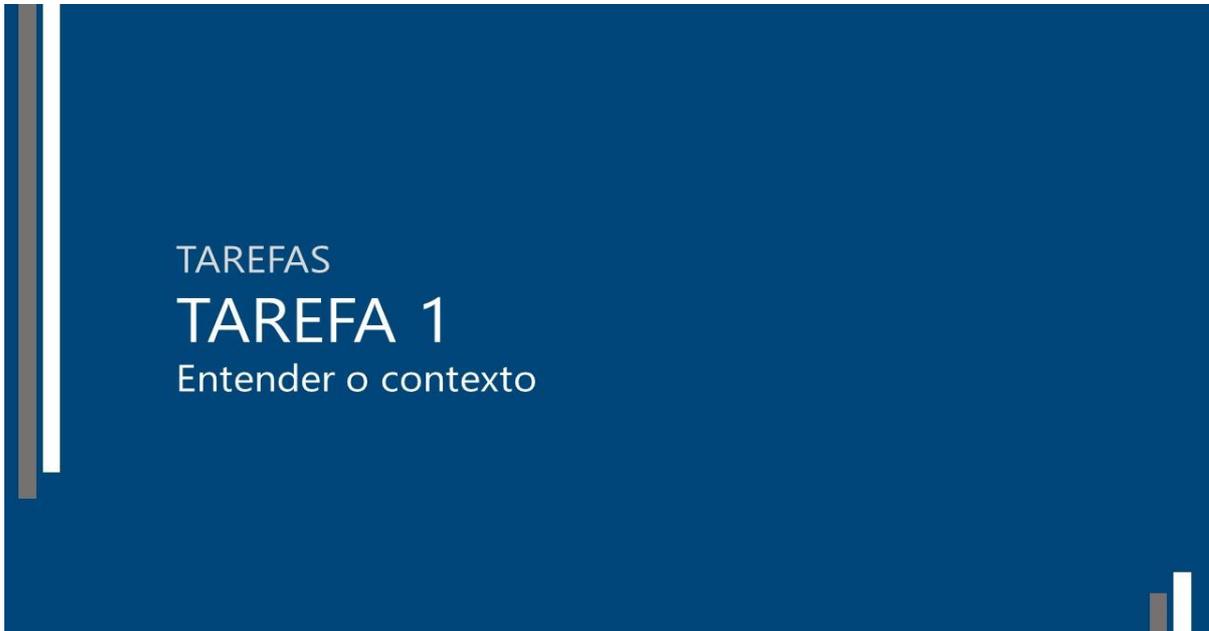
Fluxo das Tarefas

Método está composto por um conjunto de 7 Tarefas que se desdobram em suas correspondentes Atividades. Este *slide* apresenta a composição do Método e seu fluxo de execução. Em particular, são identificados os desvios que representam as possibilidades de iteração do Método. Estes desvios informam o ponto de saída (seta vermelha) para a Tarefa e a Atividade onde deve ser reiniciada e execução. A Atividade ou Tarefa onde deve acontecer o reinício do Método também está identificada neste fluxo. Quando a atividade ou tarefa estão identificadas por “XX”, isto significa a atividade ou tarefa de retomada do Método depende do resultado da execução em curso.



SLIDE 10

Este *slide* tem como objetivo organizar a apresentação.

**SLIDE 11****Tarefa 1 – Entender o Contexto**

Esta Tarefa tem como objetivo entender o contexto de aplicação do Método.

A slide with a white background and a blue border. On the left side, there are two vertical bars: a thin white one and a thicker blue one. At the top left, there is a grey horizontal bar with the text 'TAREFA 1'. Below it, there is a list of four activities: 'Atividade 1_ Avaliar os impactos da aplicação do Método', 'Atividade 2_ Conhecer os recursos disponíveis', 'Atividade 3_ Identificar os objetivos do Negócio', and 'Atividade 4_ Definir critérios de sucesso para o Negócio'. In the top right corner, there is a speaker icon and the number '5'.

SLIDE 12

Avaliar os impactos da aplicação do Método

A aplicação do Método na plataforma *DW/BI* pode gerar impactos em alguns componentes da mesma. Para direcionar a análise de impacto e identificar cada um dos componentes da plataforma, o Método considera o Modelo definido por Kimball.

Os diferentes componentes que devem ser avaliados são descritos nesta Atividade e tratados individualmente em *slides* posteriores. Repare que na execução das ações você pode precisar do apoio de especialistas.

Os Impactos na plataforma podem ser gerados principalmente pela introdução de 3 novos elementos:

- Novas ferramentas usadas para análise e visualização de dados;
- Novos conjuntos de dados diferentes dos consideradas até o momento no *Data Warehouse*, inclusive dados externos;
- Compatibilidade entre os processos já executados na plataforma *DW/BI* frente a novos processos com o mesmo propósito e que fazem parte do KDD. Cita-se como exemplo: a diferença na forma como são identificados os objetivos do Negócio quando se trabalha com BI e com KDD gera grandes diferenças na forma como os dados são tratados para a formação dos respectivos modelos de dados.

Na medida em que for avançando na análise dos impactos, gere uma relação dos mesmos. Finalizando esta atividade, faça uma análise e elabore um plano de contingência para cada um dos impactos encontrados.

6 

Atividade 1_ AVALIAR OS IMPACTOS DA APLICAÇÃO DO MÉTODO

Entenda o atual contexto

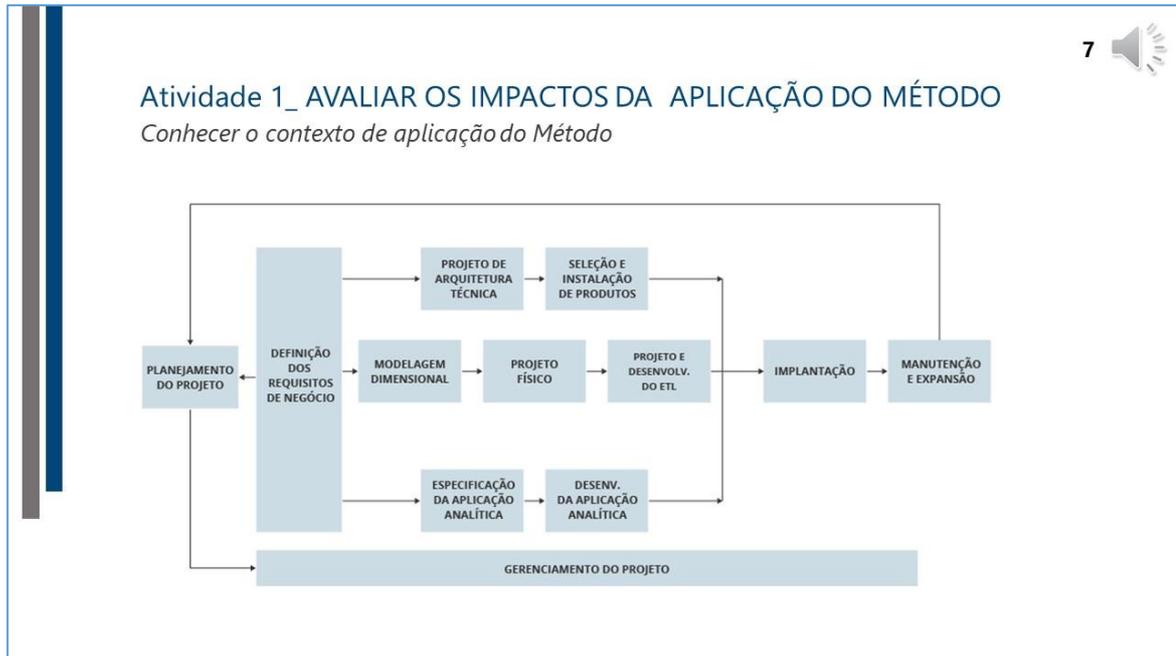
- 1_ Conhecer o contexto de aplicação do Método
- 2_ Identificar objetivos do Negócio
 Analista de Negócios
- 3_ Avaliar possíveis impactos no Projeto de Arquitetura Técnica
 Arq. Téc, Gest. de Matadados, Adm. de Dados
- 4_ Avaliar possíveis impactos na Seleção e Instalação de Produtos
 Arquiteto Técnico, Gestor de Matadados
- 5_ Avaliar possíveis impactos na Modelagem Dimensional
 Administrador de Dados, Analista de Negócios

- 6_ Avaliar possíveis impactos na Modelagem Física
 Administrador de Dados, Analista de Negócios
- 7_ Avaliar possíveis impactos no Desenvolvimento do ETL
 Arquiteto de ETL
- 8_ Avaliar possíveis impactos na implantação.
 Analista de Negócios
- 9_ Gerar uma relação de possíveis Impactos e Contingências.
 _ Relação de Impactos e Contingências

SLIDE 13

Avaliar os impactos da aplicação do Método - Conhecer o contexto da aplicação do Método

O Modelo de Kimball descreve um completo sistema *end-to-end* de *Data Warehouse* e *Business Intelligence*. O *Data Warehouse* é um conjunto de dados onde é possível trabalhar com aplicativos de *BI*. Kimball define *BI* como um termo genérico, empregado para descrever um ativo organizacional usado para reunir informação interna e externa à organização, aplicada para dar apoio à tomada de decisão. Kimball chama este binômio de *DW/BI*. Para este binômio, o autor define um modelo sob o rótulo de “ciclo de vida”. Este Modelo pode ser usado tanto para a implantação da nova tecnologia quanto para qualquer tipo de manutenção de seus componentes após feita a implantação.

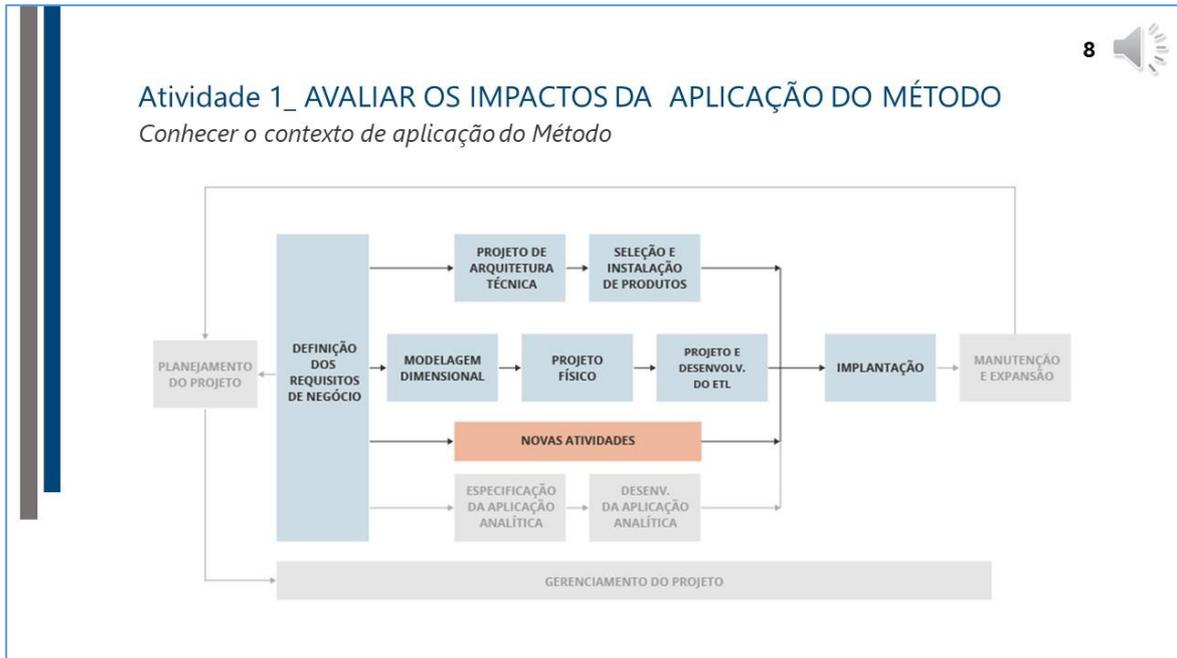


SLIDE 14

Avaliar os impactos da aplicação do Método -Conhecer o contexto da aplicação do Método

Um conjunto de novas tarefas associadas à plataforma *DW/BI* são definidas para poder executar o processo de KDD. Estas tarefas estão representadas em um quadro rosa no esquema do Modelo.

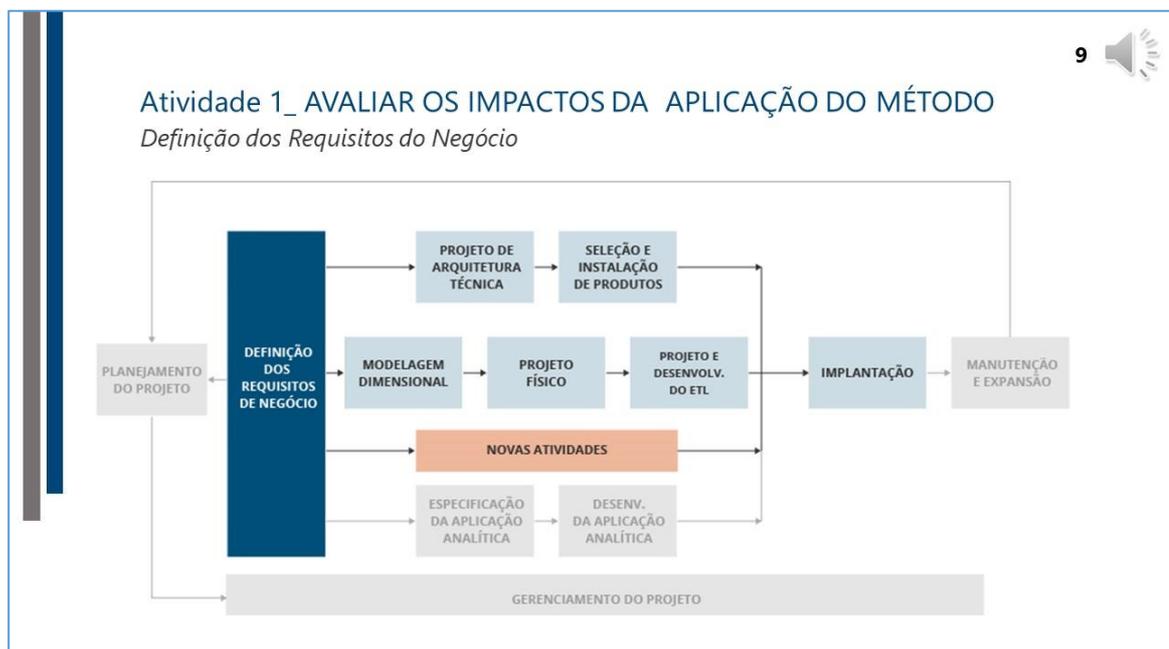
Os elementos dentro da plataforma *DW/BI* que podem ser impactados pela aplicação do Método, e consequente introdução das novas tarefas, estão representados em AZUL. A análise destes impactos será realizada a seguir.



SLIDE 15

Avaliar os impactos da aplicação do Método - Identificar os Requisitos do Negócio

Da mesma forma como acontece no sistema de *BI*, também são os objetivos do Negócio que direcionam todo o trabalho de KDD. Porém, existe uma grande diferença na forma em como são identificados os objetivos em ambos contextos.



SLIDE 16**Avaliar os impactos da aplicação do Método - Identificar os Objetivos do Negócio**

Em um sistema de *BI* as hipóteses são levantadas pelo Negócio. Espera-se que os dados irão validar, ou não, estas hipóteses. No processo de KDD, o Negócio levanta questionamentos para os quais os dados informam alguns padrões que posteriormente serão validados pelo Negócio.

O correto entendimento de como esta atividade deve ser executada é fundamental, já que o resultado da sua execução se converte em insumos para todas as tarefas que serão executadas posteriormente. Vejamos um exemplo de aplicação: a área de marketing de uma organização pretende trabalhar a fidelização de clientes. Em tal caso, a área solicita ao Analista de *BI* uma análise por idade, local de residência e renda dos clientes catalogados como fiéis. Desta forma, pretende avaliar se estas informações têm influência na fidelização. Já no processo de KDD, o Negócio apresenta como requisito um questionamento: como fidelizar clientes?

Atividade 1_ AVALIAR OS IMPACTOS DA APLICAÇÃO DO MÉTODO

Identificar os objetivos do Negócio

Para um sistema de BI

- _ Hipóteses definidas/subentendidas pelo negócio;
- _ Procura relações de Causa e Efeito;
- _ Faz uso de referências e experiências anteriores;
- _ Analisa dados contextualizados, relacionados.

10 

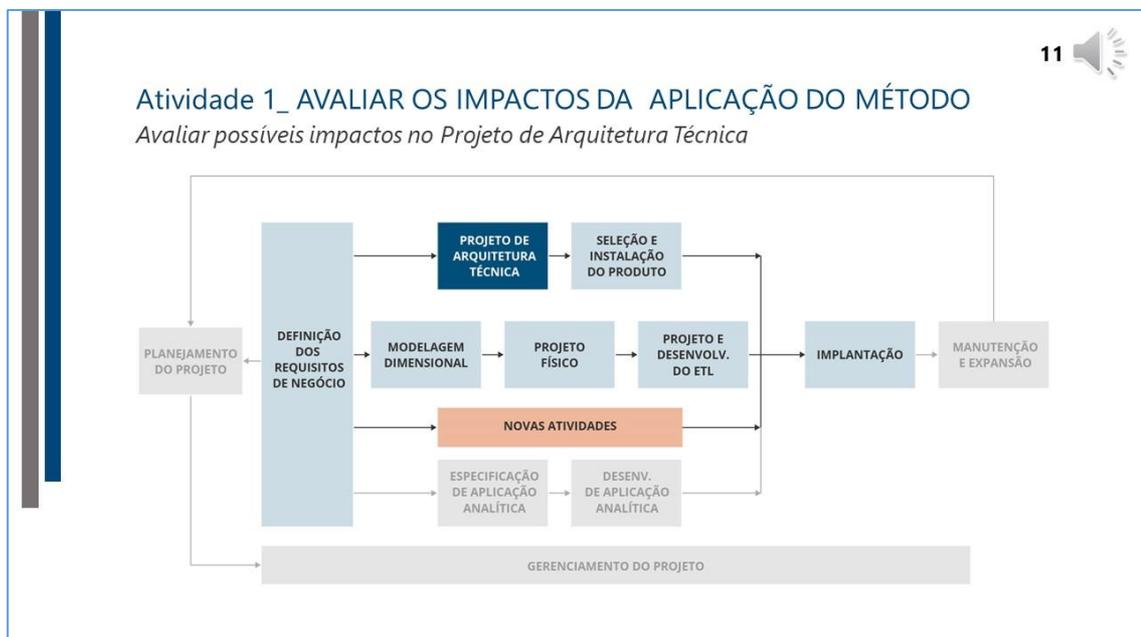


Para um sistema de KDD

- _ O Negócio levanta questionamentos;
- _ Os dados geram padrões potencialmente úteis;
- _ O negocio valida os conhecimento gerado.

SLIDE 17**A Avaliar os impactos da aplicação do Método - Projeto de Arquitetura Técnica**

Enquanto o componente do modelo de Kimball titulado “Definição dos Objetivos do Negócio” define “o que fazer”, o componente “Projeto da Arquitetura Técnica” define “como fazer”. O projeto de Arquitetura Técnica mapeia todo o fluxo de dados desde sua fonte até serem disponibilizados para o usuário, além de definir como será realizada sua armazenagem.

**SLIDE 18****Avaliar os impactos da aplicação do Método - Projeto de Arquitetura Técnica**

A medida em que for realizando a avaliação dos itens apresentados, crie uma relação de possíveis impactos conjuntamente com opções de tratamento ou contingências para os mesmos. Lembrando que os 3 elementos geradores de impacto são:

- Introdução de novas ferramentas para dar suporte ao processo de KDD;

- Novos dados, inclusive externos, poderão ser incorporados ao *Data Warehouse*;
- Os Processos de tratamento de dados realizados para o ETL e para a formação do Modelo de *BI* podem tanto colaborar quanto entrar em conflito com o processo de tratamento de dados que faz parte do processo de KDD.

A partir das questões levantadas, deve ser considerado o impacto nos seguintes componentes da arquitetura técnica:

Fontes de Dados: neste componente são reunidas todas as fontes de dados que serão usadas na plataforma. Considera tanto os dados internos quanto os externos.

Metadados: os metadados são um conjunto de informações que definem e descrevem a estrutura, procedimentos e conteúdo da plataforma *DW/BI*. Eles fornecem informação necessária para permitir que as aplicações sejam executadas. Os Metadados podem ser divididos em 3 categorias:

- **Metadados técnicos:** definem processos e objetos da plataforma *DW/BI* desde um ponto de vista técnico. Exemplos: tabelas, campos, atividades e frequência das mesmas.
- **Metadados do negócio:** descrevem o conteúdo da plataforma *DW/BI*. Exemplos: origem e nome dos campos.
- **Metadados de processos:** descrevem os resultados de diversos procedimentos executados na plataforma *DW/BI*. Exemplos: tempo de início e fim, área e número de processos.

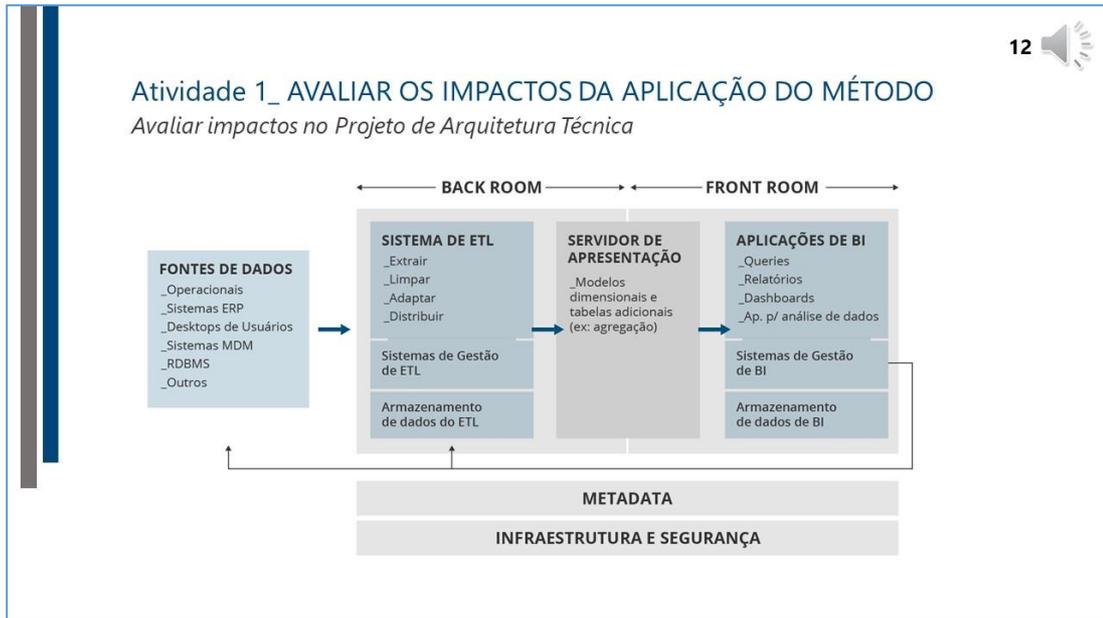
Infraestrutura e Segurança: fornece toda a base ou suporte para a arquitetura da plataforma *DW/BI*. Em particular, no momento de acrescentar novas ferramentas devem ser considerados os seguintes recursos: memória, área em disco, compartilhamento de plataformas, uso da rede e segurança no acesso.

Sistema de ETL: *Extract/Transform/Load*, processo mediante o qual os dados obtidos de diferentes fontes são preparados para a carga no *Data Warehouse*. Pode-

se dizer que, como regra, o processo de ETL consome 70% do tempo investido no tratamento de dados até eles ficarem disponíveis para o usuário final. Este é um dado extremamente importante quando se tratar de avaliar a criação do Modelo de Dados durante o processo de KDD. Estes modelos podem ser, dependendo dos requisitos do negócio, apenas temporários. Simplificar o processo de ETL pode ser uma boa opção. Também, o tratamento nos dados que acompanha a execução do ETL pode estar ou não alinhado com o tratamento de dados realizado no processo de KDD.

Servidor de Apresentação: é onde os dados são armazenados para posterior acesso do usuário (*queries*). O Servidor de Apresentação considera a criação de estruturas intermediárias de armazenamento que permitem disponibilizar os dados de forma que a execução de aplicativos para interação com os usuários apresente melhor performance. Dependendo da necessidade do Negócio, os dados para o processo de KDD podem ser diretamente extraídos desta área de armazenamento.

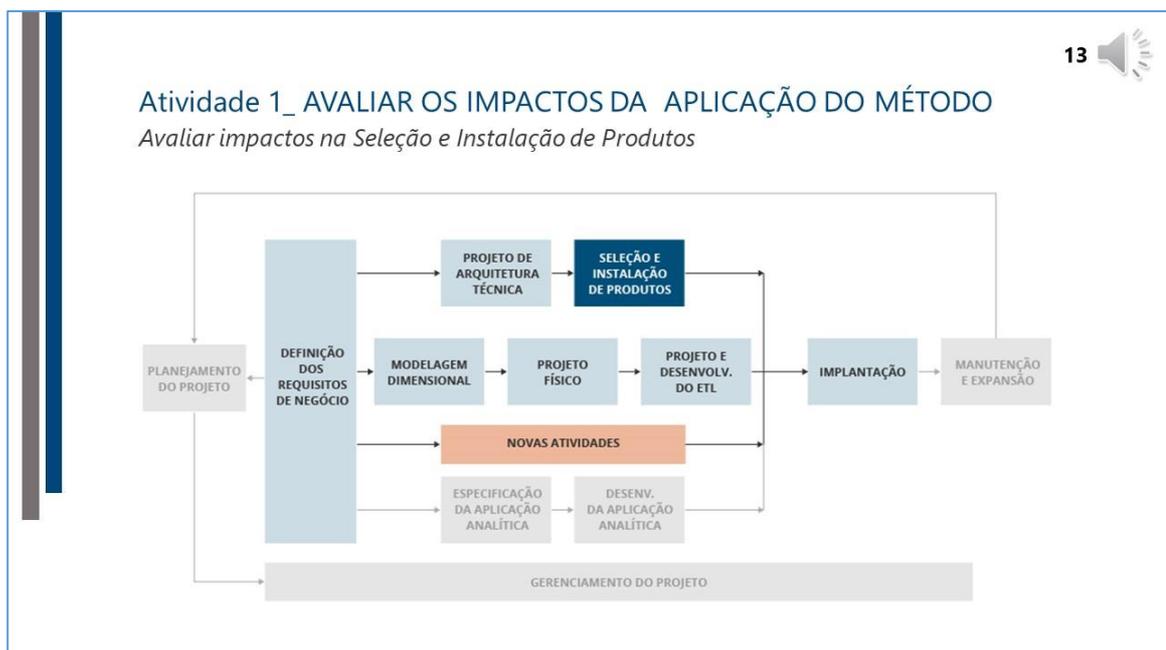
Aplicações de BI: representam a interface entre o usuário e a plataforma *DW/BI*. É nesta estrutura onde serão inseridas as ferramentas que darão suporte ao processo de KDD. Por este motivo, precisam ser avaliados os seguintes elementos: acesso à internet, segurança de dados, metadados, acesso a portais, área de armazenamento temporário para execução das novas transações, entre outros. Cabe destacar o componente de armazenamento temporário para análise de dados chamado "Armazém de Dados Analíticos Temporários". Este é usado para guardar dados de curto tempo de uso ou dados parciais de um trabalho. Esta área é de extrema importância para o processo de tratamento de dados realizado durante o processo de KDD, certamente receberá um grande volume de dados.



SLIDE 19

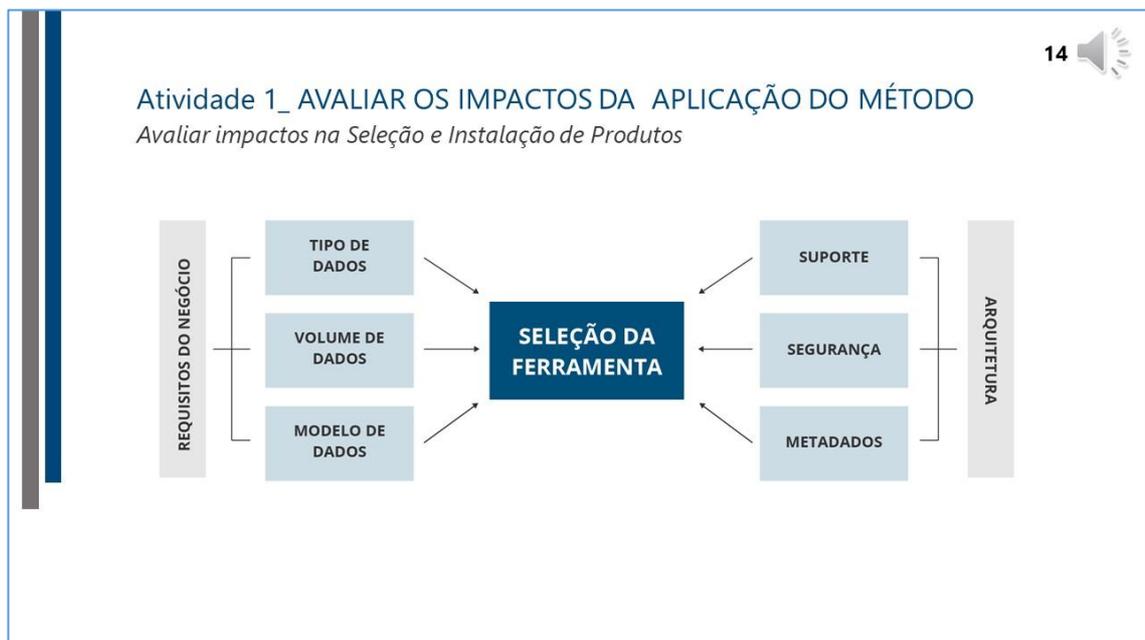
Avaliar os impactos da aplicação do Método - Seleção e Instalação de Produtos

O componente “Seleção e instalação dos Produtos” trata do processo de seleção e instalação das ferramentas de Software que serão usadas na plataforma DW/BI.



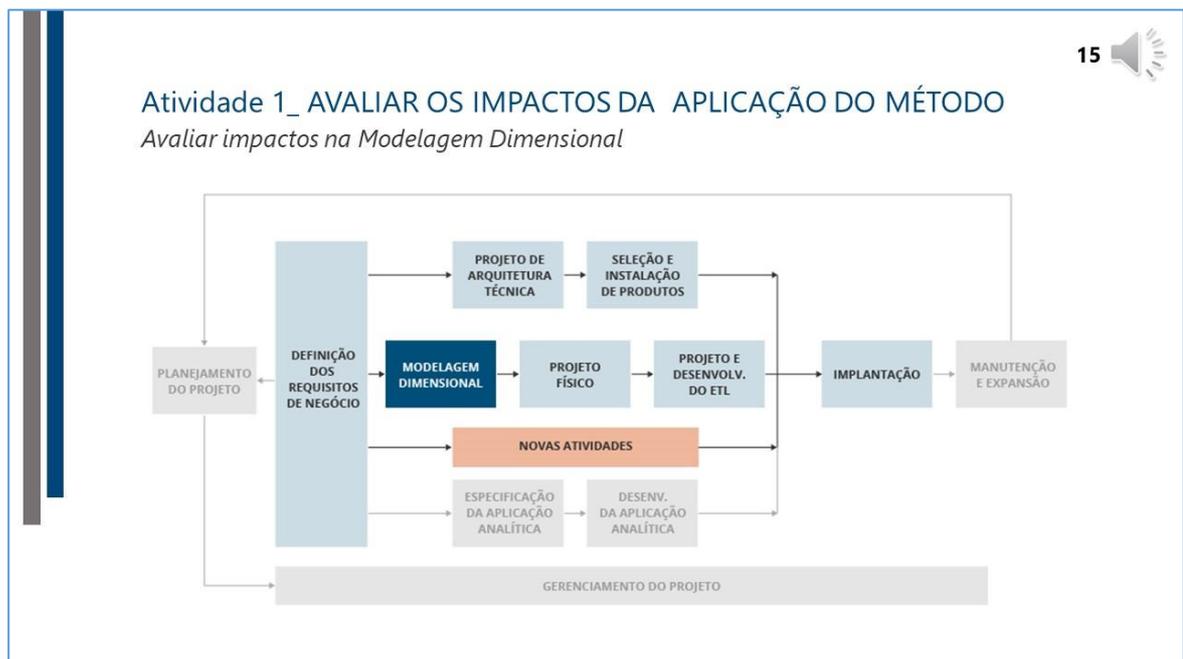
SLIDE 20**Avaliar os impactos da aplicação do Método - Seleção e Instalação dos Produtos**

A seleção inicial de ferramentas para o processo de KDD está associada aos requisitos do Negócio e à capacidade de instalação dentro da arquitetura da plataforma *DW/BI*. Durante a execução do processo de KDD, os requisitos do Negócio são transformados em requisitos para a formação do Modelo de Dados. Também devem ser consideradas outras variáveis como: volume e tipo de dados. Somente em tempo de definição do Modelo de Dados será possível identificar claramente a ferramenta que será usada para a construção do mesmo.

**SLIDE 21****Avaliar os impactos da aplicação do Método - Modelagem Dimensional**

A formação do Modelo Dimensional de *BI* deve ser tratada com muito cuidado, já que o tratamento de dados para a elaboração do Modelo Dimensional pode tanto facilitar e apoiar quanto entrar em conflito com o tratamento de dados realizado durante o processo de KDD. Vejamos alguns exemplos:

- Exemplo 1: para trabalhar com a Modelagem Dimensional, o Analista de BI pode decidir excluir campos que tenham maior parte do seu conteúdo nulo. Trabalhando com o processo de KDD, você pode aplicar algoritmos de Inteligência Artificial para poder preencher estes campos.
- Exemplo 2: dados identificados como “ruído”, dados fora do domínio, são normalmente eliminados quando se faz a Modelagem Dimensional. Porém, para o processo de KDD, estas exceções podem trazer padrões diferenciados que agreguem conhecimento de valor ao Negócio. Um exemplo desta colocação são os dados fora do domínio encontrados durante a detecção de fraudes sobre sistemas de gestão de seguros.

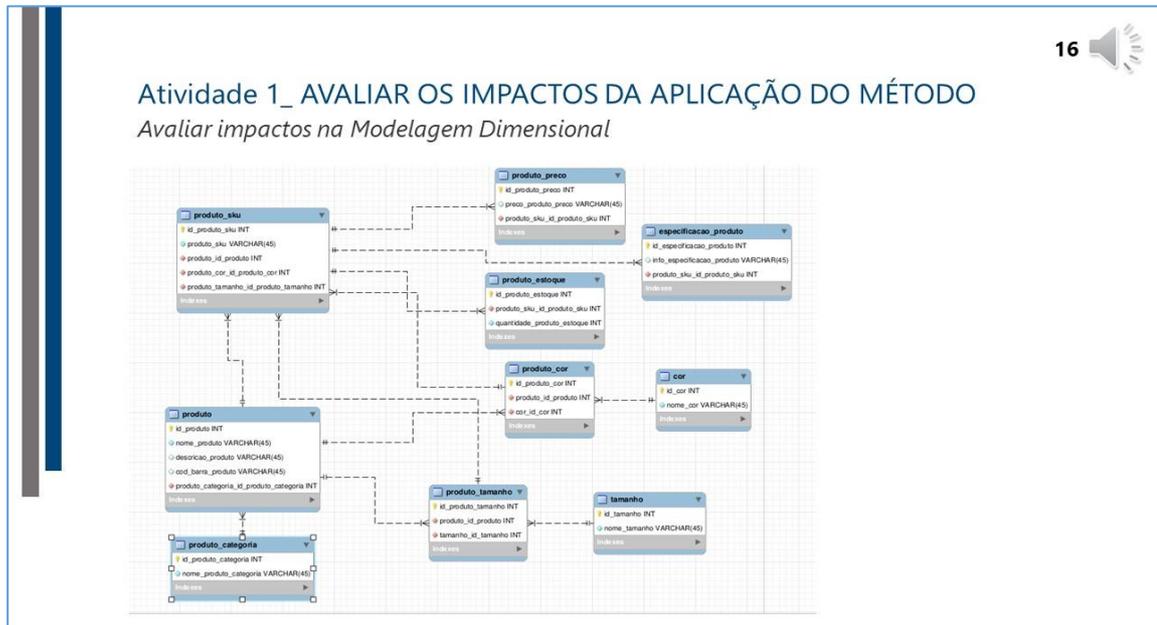


SLIDE 22

Avaliar os impactos da aplicação do Método - Modelagem Dimensional

Um Modelo Dimensional é uma forma de estruturar os dados de forma que estes sejam apresentados em forma intuitiva ao Negócio. Elaborar um Modelo Dimensional pode requerer de diversas interações com o Negócio até obter um

modelo sólido. Dependendo da complexidade, pode demandar um longo período para ficar pronto.



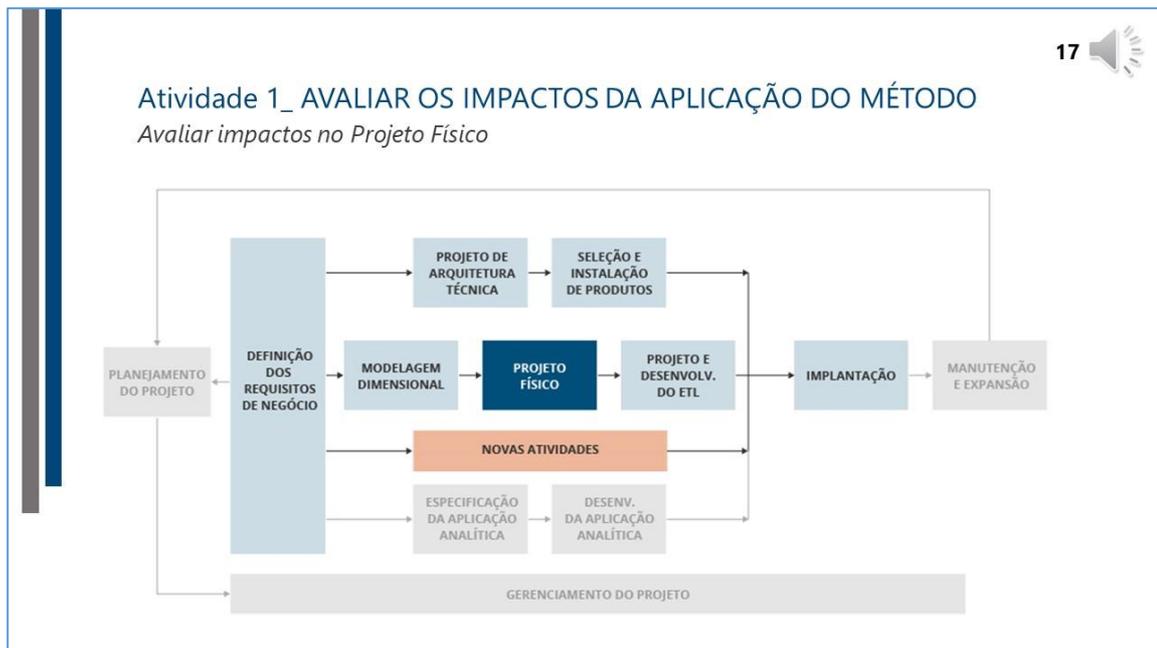
Fonte:

https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKewjqnpWj_LcAhVBvJAKHSHwDAQQiRx6BBAgBEAU&url=https%3A%2F%2Fpt.stackoverflow.com%2Fquestions%2F139902%2Fmodelo-relacional-banco-de-dados-com-sku&psig=AOvVaw3CE193ps9CLc1HrZ37iEBW&ust=1534527664804704, nov/2018

SLIDE 23

Avaliar os impactos da aplicação do Método - Projeto Físico

O Desenho Físico dos dados transforma o Modelo Lógico de dados em uma estrutura física no *Data Warehouse*.



SLIDE 24

Avaliar os impactos da aplicação do Método - Projeto Físico

O Desenho Físico deve considerar outros requisitos, como as ferramentas de acesso e performance do *query* nos dados. Deverão ser definidas as tabelas, índices, particionamento e agregação, entre outros elementos. Lembre-se que a Construção do Modelo Físico surge a partir da definição do Modelo Dimensional e gera requisitos para a elaboração o do ETL.

Atividade 1_ AVALIAR OS IMPACTOS DA APLICAÇÃO DO MÉTODO

Avaliar impactos no Desenho Físico



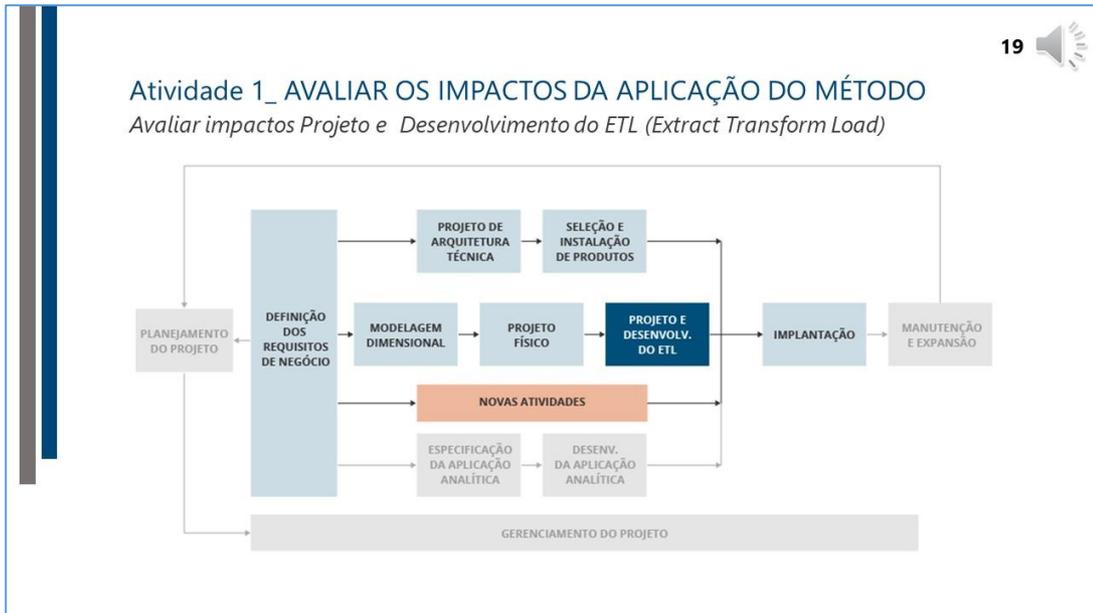
Fonte:

https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKewjgnpyWi_LcAhVBvJAKHSHwDAQjRx6BAGBEAU&url=https%3A%2F%2Fandersoncampos.wordpress.com%2F2012%2F01%2F19%2Fmodelagem-dimensional-datamart-de-vendas%2F&psig=A0Vvaw3CE193ps9CLc1HrZ37iEBW&ust=1534527664804704, nov/2018

SLIDE 25

Avaliar os impactos da aplicação do Método - Projeto e Desenvolvimento do ETL

A construção do ETL está determinada por diferentes requisitos que precisam ser considerados e integrados. Este fato torna esta atividade uma atividade complexa. Entre outros elementos, devem-se considerar: requisitos do negócio, fontes de dados, orçamento e janelas de processamento. Existem 34 subsistemas ou procedimentos que compõem o ETL, fato este que comprova a complexidade do processo.



SLIDE 26

Avaliar os impactos da aplicação do Método – Projeto e Desenvolvimento do ETL

Entre os 34 subsistemas e procedimentos que compõem o ETL, apenas se mencionam aqueles que podem-se ver impactados pela aplicação do Método.

Necessidade do Negócio: basicamente conduz todo o processo de ETL, lembrando que a necessidade do Negócio é identificada de forma diferenciada para ao processo de KDD.

Qualidade dos Dados: o tratamento de dados deve estar alinhado com o processo de tratamento de dados, que faz parte do processo de KDD.

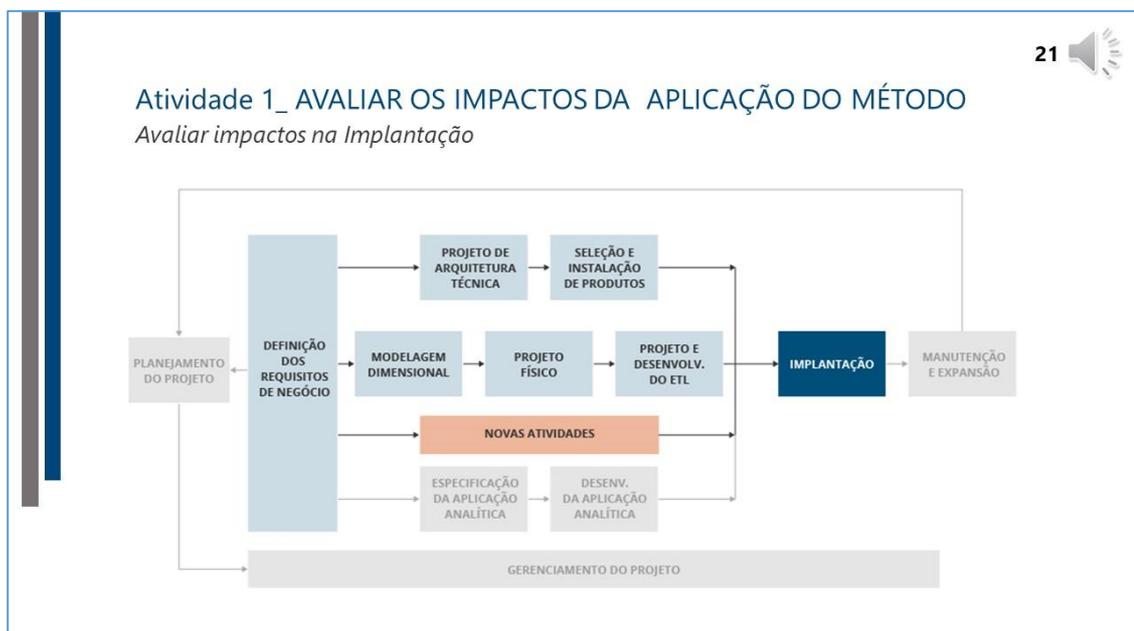
Integração de dados: lembre-se que novas fontes de dados poderão ser consideradas para o processo de KDD, inclusive as externas. Portanto, a forma como se fará a integração dos dados pode ser afetada.



SLIDE 27

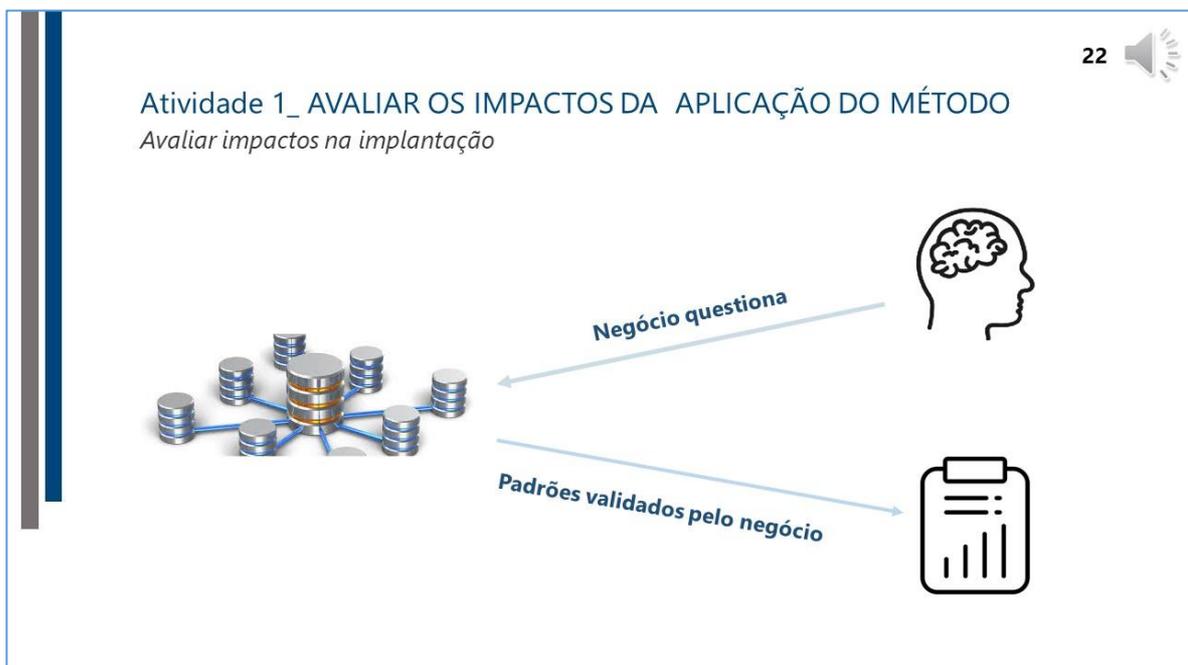
Avaliar os impactos da aplicação do Método - Implantação

Durante a Implantação, os resultados são apresentados ao Negócio. Considerando a forma diferenciada como os requisitos são identificados, a forma como os resultados são apresentados e validados pelo Negócio também é diferente.



SLIDE 28**Avaliar os impactos da aplicação do Método - Implantação**

Iniciando o processo, o Negócio levanta questionamentos que os dados devem responder com a geração de Padrões. Os Padrões apresentados ao Negócio devem ser validados por especialistas do domínio de conhecimento. Na validação, consideram-se 4 quesitos que devem ser cumpridos pelos Padrões, estes devem ser: compreensíveis, válidos, novos e úteis.

**SLIDE 29****Conhecer os Recursos Disponíveis – Entenda o atual Contexto**

Essa Atividade envolve a descoberta dos recursos necessários para a execução do Método. Executar esta atividade é fundamental para entender as restrições que podem existir para aplicação do Método e conseqüentemente para avaliar a viabilidade de execução do processo KDD.

Ações

1. Verifique se o conhecimento do usuário do Método está de acordo com as competências identificadas anteriormente, durante a descrição do funcionamento do Método. Caso não possua todas as competências necessárias, sugere-se procurar o apoio de especialistas, identificados em cada uma das ações que compõem as atividades;
2. Conheça sua plataforma *DW/BI*. Se possível, vá além dos itens avaliados durante o estudo de impactos da aplicação do Método;
3. Faça um inventário das ferramentas da sua plataforma destinadas à visualização e análise de dados;
4. Verifique a possibilidade de ter recursos para compra de novas ferramentas;
5. Faça um inventário das fontes de dados disponíveis;
6. Verifique a possibilidade obter novos dados externos. Valide a disponibilidade dos dados, questões legais e custos envolvidos;
7. A partir das ações executadas, crie uma **Relação dos Recursos** disponíveis e indique quais recursos podem ser considerados como premissas (suposições certas) e quais devem ser considerados como restrições (imposições condicionantes) para a execução deste Método.

23 

Atividade 2_ CONHECER OS RECURSOS DISPONÍVEIS

Entenda o atual contexto

- 1_ Identificar o Conhecimento do Usuário do Método
 Arquiteto Técnico, Gestor de Metadados, Arquiteto de ETL
- 2_ Conhecer a plataforma DW/BI
- 3_ Fazer um inventário de ferramentas de visualização e análise de dados disponíveis
- 4_ Avaliar a possibilidade de adquirir novas ferramentas
 Analista de Dados
- 5_ Fazer um inventário das fontes de dados disponíveis
 Analista de Dados
- 6_ Verificar a possibilidade de poder obter novos dados de fontes externas
- 7_ Gerar uma relação dos **Recursos Disponíveis**

 _ Impactos e Contingências

 _ Recursos Disponíveis

SLIDE 30

Identificar os Objetivos do Negócio

O objetivo desta atividade é entender profundamente o ponto de vista do Negócio, ou seja, o que o cliente realmente quer realizar. Muitas vezes, o cliente tem muitos objetivos e restrições concorrentes que devem ser adequadamente equilibrados. Deve-se descobrir no início do projeto fatores importantes que possam influenciar o resultado. Uma consequência provável de negligenciar esta etapa seria gastar um grande esforço produzindo as respostas corretas para as perguntas erradas.

Ações:

1. Identificar a área do Negócio (por exemplo, marketing, atendimento ao cliente, desenvolvimento de negócios, etc.);
2. Identificar pessoas chave na empresa e suas funções;
3. Identificar outras unidades de negócios que são afetadas pelos objetivos do Negócio (por exemplo: novo conhecimento identificado na área de Marketing pode gerar impactos para Vendas e Finanças);
4. Definir como identificar os objetivos. Se for preciso, procure uma metodologia específica. É interessante envolver todas as áreas afetadas;
5. Especificar os benefícios esperados em termos do Negócio. Eles devem ser atingíveis e mensuráveis;
6. Documentar as informações obtidas durante a execução das ações anteriores, criando o relatório de **Objetivos do Negócio**.

24 

Atividade 3_ IDENTIFICAR OS OBJETIVOS DO NEGÓCIO

- 1_ Identificar a área do Negócio;
 Analista de Negócios
- 2_ Identificar as pessoas-chave para mapear os objetivos
 Analista de Negócios
- 3_ Identificar as unidades de negócio que são afetadas com os objetivos do processo de KDD
 Analista de Negócios
- 4_ Definir como identificar os objetivos
 Analista de Negócios
- 5_ Especificar os benefícios esperados pelo negócio
 Analista de Negócios
- 6_ Documentar os **Objetivos do Negócio**

 _Objetivos do Negócio

SLIDE 31

Definir os Critérios de Sucesso para o Negócio

Descreva os critérios para um resultado bem-sucedido ou útil para o projeto do ponto de vista do Negócio.

Ações

1. Especifique os critérios de sucesso de forma que estes sejam atingíveis e mensuráveis. Cada um dos objetivos do negócio deve ter associado ao menos um critério de sucesso;
2. Identifique quem fará a avaliação necessária uma vez finalizado o processo de KDD;
3. Documente estes critérios no documento usado para definir os **Objetivos do Negócio**;
4. Se não for possível identificar critérios de sucesso, volte à Tarefa 1, Atividade 3 – Identificar Objetivos do Negócio.

25 

Atividade 4_ DEFINIR OS CRITÉRIOS DE SUCESSO PARA O NEGÓCIO

- 1_ Especificar os critérios de sucesso para o negócio (atingíveis e mensuráveis)
 Analista de Negócios
- 2_ Indicar quem avalia os critérios de sucesso
 Analista de Negócios
- 3_ Incrementar a documentação com os **Objetivos do Negócio**
- 4_ Se os Critérios de Sucesso para o Negócio não forem viáveis, volte à TAREFA 1/Atividade 3 _ IDENTIFICAR OBJETIVOS DO NEGÓCIO


 _ Objetivos do Negócio

 _ Objetivos do Negócio

SLIDE 32

Este *slide* tem como objetivo organizar a apresentação.

TAREFAS
TAREFA 2
Entender o Modelo de Dados

SLIDE 33**Tarefa 2 – Entender o Modelo de Dados**

Esta Tarefa tem como objetivo identificar os requisitos para construção e avaliação do Modelo de Dados, assim como avaliar a sua viabilidade. Esta tarefa pode ser entendida como a transformação da Necessidade do Negócio em requisitos de dados com a expectativa de que o processo de KDD aplicado nestes dados atenderá a necessidade do Negócio.

26 

TAREFA 2

- Atividade 1_ Identificar os requisitos para a construção do Modelo de Dados
- Atividade 2_ Avaliar as premissas para a construção do Modelo de Dados
- Atividade 3_ Avaliar as restrições para a construção do Modelo de Dados
- Atividade 4_ Avaliar a viabilidade para a Construção do Modelo de Dados
- Atividade 5_ Determinar os objetivos da aplicação do Modelo de Dados
- Atividade 6_ Identificar os critérios de sucesso da aplicação do Modelo de Dados
- Atividade 7_ Fazer uma avaliação inicial das ferramentas para construir o Modelo de Dados



- _ Recursos Disponíveis
- _ Objetivos do Negócio



- _ Especificações do Modelo de Dados
- _ Relação de ferramentas

SLIDE 34**Identificar os Requisitos para a construção do Modelo de Dados**

Esta Atividade tem como objetivo identificar os requisitos para a construção do Modelo de Dados. Entende-se como Modelo de Dados a habilidade de aplicar algoritmos a um conjunto de dados para prever atributos de destino.

Ações:

1. Identificar o usuário alvo do negócio de quem você obterá as informações de que precisa;
2. Capturar os requisitos do modelo: precisão, capacidade de implementação, capacidade de manutenção e repetibilidade do projeto de mineração de dados e dos modelos resultantes;
3. Capturar requisitos adicionais: segurança, restrições legais, privacidade;
4. Gerar um registro com estas informações. Você irá acrescentado ainda mais informações na medida que forem executando as atividades dentro desta tarefa. Este registro será identificado como **Especificações do Modelo de Dados**.

27 

ATIVIDADE 1_ IDENTIFICAR OS REQUISITOS PARA A CONSTRUÇÃO DO MODELO DE DADOS

- 1_ Identificar o usuário alvo  Analista de Negócios
- 2_ Captar os requisitos do modelo  Analista de Negócios
 - _ Precisão
 - _ Capacidade de manutenção
 - _ Repetibilidade
- 3_ Captar outros requisitos adicionais  Analista de Negócios
 - _ Segurança
 - _ Restrições legais
 - _ Privacidade
- 4_ Gerar a Relação de **Especificações do Modelo de Dados**

 _ Objetivos do Negócio

 _ Especificações do Modelo de Dados

SLIDE 35**Avaliar as Premissas para a construção do Modelo de Dados.**

Na execução desta atividade, você fará uma análise mais profunda, tendo como referência a análise preliminar realizada durante a execução da Tarefa 1.

Ações:

1. Esclarecer todas as suposições possíveis vinculadas ao conteúdo de dados questionando junto ao usuário chave e tornando-as explícitas. Use como base os objetivos do Negócio já identificados;
2. Esclarecer todas as suposições vinculadas aos dados. Exemplo: disponibilidade dos dados, volume e intervalo de tempo para formação de histórico;
3. Esclarecer as suposições sobre fatores externos. Exemplo: questões econômicas, produtos competitivos e avanços técnicos;
4. Identificar como os resultados da aplicação do Modelo serão apresentados ao solicitante;
5. Incrementar com as informações resultantes da execução desta atividade o registro de **Especificações do Modelo de Dados** iniciado na Atividade anterior.

28 

ATIVIDADE 2_ AVALIAR AS PREMISSAS PARA A CONSTRUÇÃO DO MODELO DE DADOS

- 1_ Esclarecer todas as suposições vinculadas ao conteúdo do conjunto de dados
 Analista de Negócios
- 2_ Esclarecer todas as suposições vinculadas aos dados
 Analista de Negócios, Analista de Dados
- 3_ Esclarecer suposições e restrições em relação a fatores externos
 Analista de Negócios
- 4_ Identificar como os resultados da aplicação do Modelo serão apresentados ao solicitante
 Analista de Negócios
- 5_ Incrementar a Relação de **Especificações do Modelo de Dados**

 _ Recursos Disponíveis;
 _ Objetivos do Negócio;

 _ Especificações do Modelo de Dados

SLIDE 36**Avaliar as restrições para a construção do Modelo de Dados**

Ainda com a visão de fazer uma análise mais profunda sobre o Modelo de Dados, faça uma avaliação das restrições para sua construção.

Ações:

1. Verificar as restrições gerais. Exemplo: questões legais, orçamento, cronogramas e recursos;
2. Verificar os direitos de acesso às fontes de dados. Exemplo: restrições de acesso e senha obrigatória;
3. Verificar acessibilidade técnica dos dados. Exemplo: sistemas operacionais, sistema de gerenciamento de dados, arquivo ou formato do banco de dados.
4. Verifique as restrições orçamentárias. Exemplo: custos fixos e custos de implementação.
5. Incremente com estas informações a relação de **Especificações do Modelo de Dados**.

29 

Atividade 3_ AVALIAR AS RESTRIÇÕES PARA A CONSTRUÇÃO DO MODELO DE DADOS

- 1_ Verificar as restrições gerais
 Analista de Negócios
- 2_ Verificar os direitos de acesso às fontes de dados
 Analista de Negócios, Analista de Dados
- 3_ Verificar acessibilidade técnica de dados
 Arquiteto Técnico
- 4_ Verificar as restrições orçamentárias
 Analista de Negócios
- 5_ Incrementar a Relação de **Especificações do Modelo de Dados**



- _ Especificações do Modelo de Dados;
- _ Recursos Disponíveis;
- _ Objetivos do Negócio;



- _ Especificações do Modelo de Dados

SLIDE 37**Avaliar viabilidade para construção do Modelo de Dados**

Tendo como base as especificações já identificadas para o Modelo de Dados, principalmente nas premissas e restrições, pergunte-se:

- É possível trabalhar com as restrições impostas?
- As premissas são válidas?

- Efetivamente contará com as premissas? O que pode faltar?
- É possível contornar as restrições impostas?

Se não for viável a construção do Modelo, aconselha-se reiniciar o processo avaliando novamente as necessidades do negócio. Volte a executar a Tarefa1, Atividade 3 – Identificar os Objetivos do Negócio.

30 

Atividade 4_ AVALIAR VIABILIDADE PARA A CONSTRUÇÃO DO MODELO DE DADOS

- 1_ Avaliar a viabilidade de trabalhar com as restrições impostas
- 2_ Avaliar se, efetivamente, as premissas são válidas
- 3_ Avaliar a possibilidade de contornar restrições impostas ou as premissas não válidas
- 4_ Se a construção do Modelo de Dados não for viável volte à TAREFA 1/Atividade 3, IDENTIFICAR OS OBJETIVOS DO NEGÓCIO 

 _ Especificações do Modelo de Dados

SLIDE 38

Determinar os objetivos da aplicação do Modelo de Dados

Uma vez avaliada a viabilidade da construção do Modelo de Dados, é necessário fazer a transformação dos Objetivos do Negócio em objetivos do Modelo de Dados. Em outras palavras, você precisa identificar como a aplicação do Modelo de Dados irá atender os Objetivos do Negócio. A partir dos Objetivos do Negócio, é possível identificar a técnica que será usada para a criação do Modelo. A tabela do próximo *slide* apresenta exemplos de como vincular o Técnica de Modelagem ao Objetivo do Negócio.

Ações:

1. Traduzir os objetivos do Negócio para os objetivos da aplicação do Modelo de Dados;
2. Especificar o tipo de técnica de modelagem de dados que será usada na construção do Modelo de Dados;
3. Incrementar a Relação de **Especificações do Modelo de Dados** com estas informações.

31 

Atividade 5_ DETERMINAR OS OBJETIVOS DA APLICAÇÃO DO MODELO DE DADOS

- 1_ Traduzir os objetivos do negócio para os objetivos da aplicação do Modelo de Dados
 Analista de Negócios
- 2_ Especificar o tipo de técnica de modelagem de dados que será usada na construção do Modelo de Dados
 Especialista Mineração
- 3_ Incrementar a Relação de **Especificações do Modelo de Dados**

 - Especificações Modelo de Dados
 - Objetivos do Negócio

 - Especificações Modelo de Dados

SLIDE 39

Objetivos do Negócio e Técnicas de Modelagem de Dados		
Objetivo Negócio	Exemplo	Técnica
Dividir registros em Classes predefinidas	Dividir Clientes em: _Apresentam e não apresentam riscos _Leais e não leais	CLASSIFICAÇÃO
Estimar valores de uma variável continua	Estimar o lucro CPF 90000-000	ESTIMATIVA
Classificar registros em classes pre-definidas em relação a um "valor future"	Classificação de clientes Churn e Não Churn	PREDIÇÃO (Classificação)
Estimar o "valor future" de uma variável continua	Predição de lucros pela aceitação de uma nova regra de crédito	PREDIÇÃO (Estimativa)
Dividir registros em Classes não definidas	Dividir clientes em diferentes causas de inadimplência	CLASSIFICAÇÃO (<i>sem supervisão</i>)
Dividir registros em segmentos (Clusters)	Identificar diferentes tipos de Clientes em uma base geral de clientes	SEGMENTAÇÃO (Cluster – <i>sem supervisão</i>)

Tabela com exemplos - não exaustiva

SLIDE 40**Identificar os Critérios de Sucesso da aplicação do Modelo de Dados**

Lembre-se que você já reúne informações sobre o Modelo de Dados no relatório das **Especificações do Modelo de Dados**. Agora você precisa definir os critérios que lhe ajudarão a avaliar se o Modelo gerado atende aos requisitos identificados.

Ações

1. Definir os critérios para um resultado bem-sucedido da aplicação do Modelo de Dados em termos técnicos. Por exemplo, um determinado nível de precisão preditiva ou acurácia. Os critérios de sucesso do Modelo também podem ser subjetivos;
2. Definir *benchmarks* (práticas do mercado) como critérios de avaliação;
3. Identificar possíveis critérios de validação que sejam subjetivos. Isto deve ser feito junto ao Negócio;
4. Definir as pessoas que farão a avaliação dos critérios de sucesso;
5. Incrementar o relatório de **Especificações do Modelo de Dados**.

Na tabela do *slide* a seguir são apresentados alguns critérios técnicos de avaliação associados a técnicas de modelagem.

32 

Atividade 6_ IDENTIFICAR OS CRITÉRIOS DE SUCESSO DA APLICAÇÃO DO MODELO DE DADOS

- 1_ Especificar os critérios de avaliação do Modelo de Dados
 Analista de Negócios
- 2_ Definir benchmarks para os critérios de avaliação
 Analista de Negócios
- 3_ Identificar possíveis critérios de validação subjetivos
 Analista de Negócios
- 4_ Definir as pessoas que farão a avaliação
 Analista de Negócios
- 5_ Incrementar a relação de **Especificações do Modelo de Dados**


_ Especificações do Modelo de Dados;
_ Objetivos do negócio


_ Especificações do Modelo de Dados

SLIDE 41

Critérios Técnicos de Sucesso conforme as Técnicas de Modelagem	
Técnica de Modelagem	Critérios de Sucesso
Classificação	Precisão, Lucro e Perda, Elevação, Simplicidade, Estabilidade, Velocidade, Tempo de Treinamento, Uso da Memória
Predição	Erro Quadrático Médio, Variação (Desvio Padrão), Simplicidade, Estabilidade, Velocidade, Tempo de Treinamento, Memória, Uso
Segmentação (Cluster)	Cluster normalizado significa Vetores de Importância Variável, Utilidade

Tabela com exemplos - não exaustiva

SLIDE 42**Fazer uma avaliação inicial das ferramentas para construir o Modelo de Dados**

Nesta Atividade é importante fazer uma pré-seleção de ferramentas para a construção do Modelo de Dados. Se até o momento todos os requisitos para a seleção da ferramenta não estão disponíveis, você pode fazer uma análise inicial.

Ações

1. Identificar ferramentas de acordo às técnicas que serão usadas atendente às informações que você já tem sobre o Modelo de Dados;
2. Considerar as ferramentas que já estão disponíveis na sua plataforma de *DW/BI*. Lembre-se que você já fez este levantamento anteriormente;
3. Escolher ferramentas e técnicas potenciais;
4. Validar os requisitos da HW das ferramentas e veja se é possível colocá-las na sua plataforma;

5. Crie uma relação das **Ferramentas Potenciais**, que será depurada quando você tiver informações mais detalhadas, que lhe ajudará a definir os requisitos com maior precisão.

33 

Atividade 7_ FAZER UMA AVALIAÇÃO INICIAL DAS FERRAMENTAS PARA CONSTRUIR O MODELO DE DADOS

- 1_ Identificar ferramentas de acordo às técnicas de modelagem que serão usadas
 Especialista em Mineração
- 2_ Considerar as ferramentas que já estão disponíveis na sua plataforma de DW/BI
 Especialista em Mineração
- 3_ Escolher ferramentas e técnicas potenciais
 Arquiteto Técnico
- 4_ Validar os requisitos de HW das ferramentas e ver se é possível colocá-las na plataforma DW/BI
- 5_ Crie uma Relação de **Ferramentas Potencias**



- _ Recursos disponíveis;
- _ Especificações do Modelo de Dados;
- _ Impactos e Contingências.



- _ Relação de Ferramentas

SLIDE 43

O objetivo deste *slide* é organizar a apresentação.

TAREFAS

TAREFA 3

Entender os Dados

SLIDE 44**Tarefa 3 – Entender os Dados**

Esta Tarefa tem como objetivo entender os dados para sua posterior manipulação. Lembre-se que as atividades que fazem parte desta Tarefa podem fazer parte do processo de ETL ou também da modelagem de dados para formação dos modelos de dados de *BI*. Avalie os impactos.

34 

TAREFA 3

- Atividade 1_ Coletar os dados iniciais
- Atividade 2_ Entender os atributos
- Atividade 3_ Verificar as Chaves
- Atividade 4_ Avaliar a qualidade dos dados
- Atividade 5_ Reavaliar a viabilidade do Modelo e formar suposições futuras



- _ Recursos Disponíveis
- _ Especificações do Modelo de Dados
- _ Relação de Impactos e Contingências



- _ Especificações do Modelo de Dados
- _ Descrição de Dados

SLIDE 45**Coletar os Dados Iniciais**

Adquira os dados ou acesso aos mesmos. Essa coleção inicial inclui o carregamento de dados, se necessário, para o entendimento dos mesmos. Por exemplo, se você pretende usar uma ferramenta específica para entender os dados, é lógico carregar seus dados nessa ferramenta. Lembre-se de que qualquer avaliação dos dados deve ser feita não apenas das fontes de dados individuais, mas também de todos os dados resultantes da fusão das fontes de dados. Isto deve-se ao fato de

que dados coletados de diferentes fontes podem gerar problemas de qualidade quando mesclados.

Ações:

1. Planejar quais informações são necessárias para elaborar o Modelo de Dados. Você pode precisar considerar apenas determinados atributos ou informações específicas;
2. Verificar se todas as informações necessárias para resolver as metas da modelagem estão realmente disponíveis;
3. Especificar os critérios de seleção dos atributos. Pergunte-se:
 - Quais atributos são necessários para atingir os objetivos da Modelagem de dados especificadas?
 - Quais atributos foram identificados como irrelevantes?
 - Quantos atributos podemos manipular com as técnicas de modelagem escolhidas?
4. Selecionar tabelas ou arquivos de interesse;
5. Selecionar dados dentro de uma tabela ou arquivo;
6. Considerar quanto tempo um histórico deve ser usado. Exemplo: mesmo que estejam disponíveis 18 meses de dados, apenas 12 meses podem ser necessários para a modelagem;
7. Validar a necessidade de obtenção de novos dados de fontes externas;
8. Gerar um documento com estas informações sobre os dados. Chamaremos de **Descrição de dados.**

35 

Atividade 1_ COLETAR OS DADOS INICIAIS

- 1_ Planejar quais informações são necessárias para elaborar o Modelo de Dados
- 2_ Verificar se todas as informações necessárias estão realmente disponíveis
- 3_ Especificar os critérios de seleção dos atributos
- 4_ Selecionar tabelas e arquivos de interesse
- 5_ Selecionar dados dentro de uma tabela ou arquivo
- 6_ Considerar quanto tempo um histórico deve ser usado
- 7_ Validar a necessidade de obtenção de novos dados de fontes externas
- 8_ Gere um documento de **Descrição de Dados**



- _ Recursos disponíveis
- _ Especificações do Modelo de Dados



- _ Descrição de Dados

SLIDE 46

Tarefa 3 – Entender os Atributos

Entender os atributos é fundamental para validar o objetivo do Modelo de Dados. Algumas técnicas para elaboração do Modelo de Dados trabalham apenas com valores numéricos e outras apenas com valores categóricos. Nestes casos, é necessário transformar os valores numéricos em categóricos ou os categóricos em numéricos. Não existe um critério único para transformação dos dados e diversas técnicas podem ser usadas de acordo com os objetivos pretendidos. Algumas das técnicas empregadas podem ser: suavização (remove valores errados dos dados), agrupamento (agrupa valores em faixas sumarizadas), generalização (converte valores muito específicos para valores mais genéricos), normalização (colocar as variáveis em uma mesma escala) e a criação de novos atributos (gerados a partir de outros já existentes).

Ações

1. Verificar acessibilidade e disponibilidade de atributos;
2. Verificar os tipos de atributos. Exemplo: numéricos, nominais ou discretos;
3. Verificar intervalos de valores dos atributos;

4. Analisar correlações dos atributos;
5. Entender o significado de cada atributo e valor do atributo em termos do Negócio;
6. Para cada atributo, calcular estatísticas básicas. Por exemplo: média, máximo, mínimo, desvio padrão e variância. Existem ferramentas para auxiliar nesta ação;
7. Entrevistar especialistas de domínio para obter sua opinião sobre a relevância do atributo:
 - Definir como tratar os atributos ausentes;
 - Definir como tratar atributos com texto livre;
 - Definir como tratar atributos nominais;
 - Definir como tratar atributos discretos;
8. Incremente a **Descrição de Dados** com as informações geradas nesta atividade.

Lembrando que muitas destas verificações já podem estar sendo realizadas no ETL ou na formação dos Modelo de Dados de *BI*. Valide se a forma como estão sendo realizadas está alinhada com o objetivo do Modelo de Dados em construção.

36 

Atividade 2_ ENTENDER OS ATRIBUTOS

- 1_ Verificar acessibilidade e disponibilidade de atributos
- 2_ Verificar os tipos de atributos (numéricos, nominais, discretos, etc)
- 3_ Verificar intervalos de valores dos atributos
- 4_ Analisar correlações dos atributos
- 5_ Entender o significado de cada atributo e valor do atributo em termos do negócio
 Analista de Negócios
- 6_ Para cada atributo, calcular estatísticas básicas
- 7_ Entrevistar especialistas do domínio para determinar a relevância do atributo
 Analista de Negócios
- 8- Incrementar o documento de **Descrição de Dados**



- _ Relação de Impactos e Contingências
- _ Especificações do Modelo de Dados
- _ Descrição de Dados



- _ Descrição de Dados

SLIDE 47**Verificar as Chaves**

Agora, você pode realizar uma análise das chaves que existem no conjunto de dados.

Ações

1. Analisar os principais relacionamentos;
2. Verificar a quantidade de sobreposições dos principais valores de atributos nas tabelas;
3. Incrementar o documento de **Descrição de Dados** com estas novas informações;
4. Agora, você já pode revisar os objetivos do Modelo verificando se, com os dados que você dispõe, consegue atingi-los.
5. Se isso não for possível, reinicie a execução do Método na Tarefa 2, Atividade 1 – Identificar os requisitos para a do Modelo e Dados.

37 

Atividade 3_ VERIFICAR AS CHAVES

- 1_ Analisar os principais relacionamentos
- 2_ Verificar a quantidade de sobreposições dos principais valores de atributos nas tabelas
- 3- Incrementar a **Descrição de Dados**
- 4_ Revisar objetivos para a construção do Modelo de Dados
- 5_ Se os objetivos do Modelo de Dados não forem viáveis volte à TAREFA 2, Atividade 1 - IDENTIFICAR OS REQUISITOS PARA A CONSTRUÇÃO DO MODELO DE DADOS





_ Especificações do Modelo de Dados
_ Descrição de Dados



_ Especificações do Modelo de Dados
_ Descrição de Dados

SLIDE 48

Avaliar a qualidade dos dados

Examine a qualidade dos dados, abordando questões como:

- Os dados estão completos?
- Estão corretos ou contém erros?
- Se houver erros, o quão comuns eles são?
- Existem valores ausentes nos dados? Se sim, como eles são representados, onde eles ocorrem e quão comuns eles são?

Ações

1. Considerar a revisão já realizada de chaves e atributos;
2. Verificar a cobertura. Exemplo, se todos os valores possíveis estão representados;
3. Verificar se os significados dos atributos e valores contidos se encaixam;
4. Identificar atributos ausentes e campos em branco;
5. Estabelecer o significado dos dados ausentes;
6. Verificar se há atributos com valores diferentes que tenham significados semelhantes;
7. Verificar a ortografia e o formato dos valores. Exemplo: mesmo valor, mas às vezes começando com uma letra minúscula, às vezes com uma letra maiúscula;
8. Verificar se há ruídos. Lembre-se que o “ruído” pode indicar um fenômeno interessante;
9. Verificar a plausibilidade dos valores;
10. Atualizar o documento de **Descrição de Dados** com estas informações.

Use ferramentas de visualização para lhe ajudar nesta atividade.

38 

Atividade 4_ AVALIAR A QUALIDADE DOS DADOS

<p>1_ Considerar a revisão já realizada em chaves e atributos</p> <p>2_ Verificar a cobertura  Analista de Negócios</p> <p>3_ Verificar se os significados dos atributos e valores contidos se encaixam  Analista de Negócios</p> <p>4_ Identificar atributos ausentes e campos em branco</p> <p>5_ Estabelecer o significado de dados ausentes  Analista de Negócios</p>	<p>6_ Verificar se há atributos com valores diferentes com significados semelhantes  Analista de Negócios</p> <p>7_ Verificar a ortografia e o formato dos valores</p> <p>8_ Verificar se há ruídos  Analista de Negócios</p> <p>9_ Verificar a plausibilidade dos valores  Analista de Negócios</p> <p>10_ Incrementar o documento de Descrição de Dados</p>
---	--



_ Relação de Impactos e Contingências
_ Descrição de Dados



_ Descrição de Dados

SLIDE 49

Reavaliar o Modelo e formar suposições para a análise futura.

Agora que você já conhece bem os dados, pode fazer uma melhor avaliação dos objetivos do Modelo avaliando sua viabilidade e critérios de sucesso. Você poderá formular questionamentos e validar se o modelo consegue respondê-los.

Ações

1. Formular questões para avaliar a aplicabilidade do Modelo;
2. Transformar estas questões em possíveis resultados da aplicação do Modelo de Dados;
3. Realizar uma análise básica para validar estes possíveis resultados;
4. Esclarecer ou tornar mais precisos os objetivos do Modelo de dados e critérios de sucesso;
5. Com estas novas informações, atualizar a **Especificação do Modelo de Dados**;
6. Se os resultados desta análise mostram a necessidade de fazer mudanças no Modelo de Dados volte à Tarefa 2, Atividade 1 – Identificar os Requisitos para a Construção do Modelo de Dados.

39 

Atividade 5_ REAVALIAR O MODELO E FORMAR SUPOSIÇÕES PARA ANÁLISE FUTURA

- 1_ Formular questões para avaliar o Modelo de Dados
 Analista de Negócios
- 2_ Transformar as questões em possíveis resultados
- 3_ Realizar uma análise básica para avaliar estes resultados
 Analista de Negócios
- 4_ Esclarecer ou tornar mais precisos os objetivos do Modelo de Dados e seus critérios de sucesso
- 5_ Atualizar a **Especificação do Modelo de Dados**
- 6_ Se o resultados desta análise mostram a necessidade de fazer mudanças no Modelo de Dados volte à **TAREFA 2, Atividade 1 - IDENTIFICAR OS REQUISITOS PARA A CONSTRUÇÃO DO MODELO DE DADOS** 

 _ Especificações do Modelo de Dados
 _ Descrição de Dados

 _ Especificações do Modelo de Dados

SLIDE 50

O objetivo deste *slide* é organizar a apresentação.

TAREFAS

TAREFA 4

Preparar os Dados

SLIDE 51**Tarefa 4 – Preparar os Dados**

Os dados já são conhecidos. É possível preparar os dados para a construção do Modelo. Se você trabalhar sobre o modelo de *BI* já pronto, lembre-se de garantir que a preparação de dados é compatível com a preparação de dados aplicada pelo processo de KDD.

40 

TAREFA 4

Atividade 1_ Selecionar dados

Atividade 2_ Limpar dados

Atividade 3_ Construir dados

Atividade 4_ Construir atributos e atributos derivados

Atividade 5_ Integrar dados

Atividade 6_ Formatar dados



- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas
- _ Descrição de Dados
- Relação de Impactos e Contingências



- _ Descrição de Dados
- Especificações do Modelo de Dados

SLIDE 52**Selecionar Dados**

Decida quais dados serão usados para formar o Modelo. Os critérios incluem: relevância para dos objetivos do Modelo, qualidade e restrições técnicas (limites no volume de dados e tipos de dados). Muitas das restrições técnicas são geradas pelas próprias ferramentas. Com esta atividade, você está iniciando o processo de formatação de dados.

Ações:

1. Coletar dados das diferentes fontes;
2. Considerar a seleção de subconjuntos de dados diferentes, como por exemplo: apenas dados que atendem a determinadas condições;
3. Considerar o uso de técnicas de amostragem para gerar conjuntos de dados de treinamento e reduzir o tamanho do conjunto;
4. Atualizar a **Descrição de Dados** com estas informações.

41 

Atividade 1_ SELECIONAR DADOS

- 1_ Coletar dados das diferentes fontes
 Analista de Dados
- 2_ Considerar a seleção de subconjuntos de dados diferentes
- 3_ Considerar o uso de técnicas de amostragem
 Especialista em Mineração
- 4_ Incremente o documento de **Especificações do Modelo de Dados**



_ Especificações do Modelo de Dados
_ Descrição de Dados



_ Especificações do Modelo de Dados

SLIDE 53**Limpar dados**

Aumente a qualidade dos dados para o nível exigido pelas técnicas de análise selecionadas. Isso pode envolver a seleção de subconjuntos limpos dos dados, a inserção de padrões adequados ou técnicas mais ambiciosas, como a estimativa de dados ausentes por modelagem.

Ações:

1. Reconsiderar como lidar com qualquer tipo de ruído observado;
2. Corrigir, remover ou ignorar o ruído;

3. Decidir como lidar com valores especiais e seu significado. Exemplo: valores especiais também podem surgir quando os dados são truncados - como, 00 para pessoas com 100 anos de idade;
4. Incremente o documento de **Descrição de Dados**.

42 

Atividade 2_ LIMPAR DADOS

- 1_ Reconsiderar como lidar com qualquer tipo de ruído observado
 Analista de Dados
- 2_ Corrigir, remover ou ignorar o ruído
 Analista de Dados
- 3_ Decidir como lidar com valores especiais e seus significados
 Analista de Dados
- 4_ Incrementar o documento de **Descrição de Dados**



- _ Especificações do Modelo de Dados
- _ Descrição de Dados
- Relação de Impactos e Contingências



- _ Descrição de Dados

SLIDE 54

Construir Dados

Essa Atividade inclui operações de preparação de dados que precisam ser construídos, como por exemplo, a transformação de valores.

Ações

1. Verificar os mecanismos de construção disponíveis com a lista de ferramentas sugeridas;
2. Decidir se é melhor realizar a construção dentro da ferramenta ou fora dela (ou seja, o que é mais eficiente, exato, repetível);
3. Incrementar o documento de **Especificações do Modelo de Dados** com as informações geradas durante a execução desta atividade.

43 

Atividade 3_ CONSTRUIR DADOS

- 1_ Verificar mecanismos de construção disponíveis com a lista de ferramentas identificadas
 Especialista em Mineração
- 2_ Decidir se é melhor realizar a construção dentro da ferramenta ou fora dela
 Especialista em Mineração
- 3_ Incrementar o documento de **Especificações do Modelo de Dados**



- _ Descrição de Dados
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas



- _ Especificações do Modelo de Dados

SLIDE 55

Construir Atributos e Atributos Derivados

Atributos derivados são novos atributos construídos a partir de um ou mais atributos existentes no mesmo registro. Um atributo precisa ser construído por diferentes motivos:

- O conhecimento de fundo nos convence de que algum fato é importante e deve ser representado, embora não tenhamos um atributo atual para representá-lo;
- O algoritmo de modelagem em uso lida apenas com certos tipos de dados;
- A avaliação do resultado esperado do Modelo sugere que certos fatos não estão sendo cobertos.

Ações:

1. Construir ou imputar atributos ausentes:
 - Adicionar novos atributos aos dados acessados.
 - Transformar atributos. Transformações podem ser necessárias para alterar intervalos para campos simbólicos (por exemplo, idades e faixas etárias) ou campos simbólicos ("definitivamente sim", "sim", "não sabe",

"não") para valores numéricos. Ferramentas de modelagem ou algoritmos geralmente exigem este tipo de ação.

2. Incremente o documento de **Especificações do Modelo de Dados** com as informações geradas durante e execução desta atividade.

44 

Atividade 4_ CONSTRUIR ATRIBUTOS E ATRIBUTOS DERIVADOS

- 1_ Considerar os seguintes critérios para a construção de atributos:
 - _ O conhecimento de fundo nos convence de que algum fato é importante e deve ser representado, embora não tenhamos um atributo atual para representá-lo
 - _ O algoritmo de modelagem em uso lida apenas com certos tipos de dados
 - _ A avaliação do resultado esperado do Modelo sugere que certos fatos não estão sendo cobertos
 - _ Construir ou imputar atributos ausentes
 - _ Adicionar novos atributos aos dados acessados
 - _ Transformar atributos
- 2_ Incrementar o documento de **Especificações do Modelo de Dados**
 -  _ Descrição de Dados
 -  _ Especificações do Modelo de Dados
 -  _ Relação de Impactos e Contingências

 Especialista em Mineração

 _ Especificações do Modelo de Dados

SLIDE 56

Integrar dados

Esta atividade tem como objetivo combinar dados provenientes de diferentes tabelas.

Ações

1. Verifique se os recursos de integração são capazes de integrar as fontes de entrada conforme necessário;
2. Integrar fontes e armazenar resultados.
3. Se integração de dados não for possível ou não estiver de acordo com o esperado, volte a executar a Tarefa 4, Atividade 1 – Selecionar Dados.

45 

Atividade 5_ INTEGRAR DADOS

- 1_ Verificar se as ferramentas de integração são capazes de integrar as fontes de entrada conforme necessário
- 2_ Integrar fontes e armazenar resultados
- 3_ Se não for possível integrar os dados pode ser necessário fazer mudanças na Seleção de Dados. Volte à **TAREFA 4, Atividade 1 - SELECIONAR DADOS**



 _ Descrição de Dados

- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas
- _ Relação de Impactos e Contingências

 _ Dados Formatados

SLIDE 57

Este *slide* tem por objetivo organizar a apresentação.

TAREFAS

TAREFA 5

Criar o Modelo de Dados

SLIDE 58**Criar o Modelo de Dados**

Esta tarefa tem como objetivo criar o Modelo de Dados.

46 

TAREFA 5

Atividade 1_ Selecionar a técnica de Modelagem

Atividade 2_ Definir o teste do Modelo de Dados

Atividade 3_ Construir o Modelo de Dados

Atividade 4_ Avaliar tecnicamente o resultado da aplicação do Modelo de Dados



- _ Descrição de Dados
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas



- _ Especificações do Modelo de Dados

SLIDE 59**Selecionar a técnica de modelagem**

Já foi visto que, de acordo com os requisitos do Negócio, algumas técnicas de Modelagem são mais apropriadas do que outras. Agora você já tem todos os requisitos necessários para fazer a melhor escolha.

Ações

1. Considerar todo o conhecimento já adquirido sobre os dados;
2. Considerar os resultados esperados;
3. Considerar a Ferramenta. Agora você já pode identificar a ferramenta certa já que conhece os dados e a técnica de modelagem que será usada. Veja alguns exemplos de ferramentas e técnicas que elas aplicam no *slide* a seguir;
4. Com estas considerações, você já pode selecionar a técnica de modelagem.
5. Incrementar as **Especificações do Modelo de Dados**.

Se várias técnicas forem aplicadas, execute esta atividade separadamente para cada técnica de forma que você poderá criar diferentes Modelos e posteriormente validar qual é o melhor.

47 

Atividade 1_ SELECIONAR A TÉCNICA DE MODELAGEM

- 1_ Considerar todo o conhecimento já adquirido sobre os dados
- 2_ Considerar os resultados esperados
- 3_ Considerar a Ferramenta
- 4_ Selecionar a técnica de modelagem
 -  Especialista em Mineração
- 5_ Incrementar as **Especificações do Modelo de Dados**



- _ Descrição de Dados
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas



- _ Especificações do Modelo de Dados

SLIDE 60

Técnicas de Modelagem	Ferramentas					
	Orange	Tanagra	Rapid Miner	KNIME	R	Weka
K-means Clustering	Sim	Sim	Sim	Sim	Sim	Sim
Association Rule Mining	Sim	Sim	Sim	Sim	Sim	Sim
Linear Regression	Sim	Sim	Sim	Sim	Sim	Sim
Logistic Regression	Sim	Sim	Sim	Sim	Sim	Sim
Naïve Bayesian Classifiers	Sim	Sim	Sim	Sim	Sim	Sim
Decision Tree	Sim	Sim	Sim	Sim	Sim	Sim
Time Series Analysis	Sim	Não	Algumas	Sim	Sim	Sim
Text Analytics	Sim	Não	Sim	Sim	Sim	Sim

Tabela de exemplos - não exaustiva

SLIDE 61**Definir o teste do Modelo de Dados**

Antes de construir um Modelo, é necessário definir um procedimento para testar a qualidade e validade do mesmo. Por exemplo, em técnicas como a “Classificação”, é comum usar taxas de erro como medidas de qualidade. Para fazer o teste, os dados devem ser separados em 2 conjuntos, o conjunto de execução e o conjunto de teste. O modelo é construído sobre o conjunto de execução e sua qualidade estimada no conjunto de testes.

Ações

1. Verificar diferentes tipos de testes para cada uma das metas definidas para o Modelo de Dados;
2. Decidir os passos necessários para execução dos testes;
3. Preparar os dados necessários para o teste;
4. Com estas informações, incrementar as **Especificações do Modelo de Dados**.

48 

Atividade 2_ DEFINIR O TESTE DO MODELO DE DADOS

- 1_ Verificar diferentes testes para cada uma das metas definidas para o Modelo de Dados
 Especialista em Mineração
- 2_ Planejar os passos necessários
 Especialista em Mineração
- 3_ Preparar os dados necessários para o teste
 Especialista em Mineração
- 4_ Incrementar as **Especificações do Modelo de Dados**



- _ Descrição de Dados
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas



- _ Especificações do Modelo de Dados

SLIDE 62**Construir o Modelo**

Com qualquer ferramenta de modelagem, muitas vezes há um grande número de parâmetros que podem ser ajustados. Os diferentes parâmetros que configuram a ferramenta levam à formação de diferentes Modelos.

Ações

1. Definir parâmetros iniciais da ferramenta;
2. Executar a técnica selecionada no conjunto de dados de entrada para produzir o **Modelo de Dados**;
3. Avaliar o resultado, a precisão esperada, robustez e possíveis deficiências. De acordo com este resultado, pode ser preciso redefinir os parâmetros da ferramenta e recriar um novo Modelo;
4. Interpretar o resultado do Modelo.
5. Se os resultados não são satisfatórios, construa outros Modelos e compare os resultados. Volte à Tarefa 5, Atividade 3 – Construir o Modelo.

49 

Atividade 3_ CONSTRUIR O MODELO

- 1_ Definir parâmetros iniciais da ferramenta
 Especialista em Mineração
- 2_ Executar a técnica selecionada no conjunto de dados de entrada para produzir o **Modelo de Dados**
 Especialista em Mineração
- 3_ Avaliar o resultado
- 4_ Interpretar o resultado do Modelo
- 5_ Se os resultados os resultados não são satisfatórios construa outros Modelos compare os resultados. Volte a **TAREFA 5, Atividade 3 - CONSTRUIR O MODELO** 



- _ Descrição dos Dados
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas

SLIDE 63**Avaliar Tecnicamente o Modelo de Dados**

O Modelo de Dados deve agora ser avaliado para garantir que ele atenda aos critérios de sucesso para ele definidos.

Ações

1. Avaliar resultados com respeito a critérios de avaliação do Modelo de dados;
2. Interpretar os resultados em termos do Negócio (tanto quanto possível nesta fase);
3. Obter comentários dos especialistas do domínio sobre os resultados;
4. Verificar a plausibilidade do modelo, ou seja, a possibilidade de ser admissível;
5. Incrementar com estas informações o documento de **Especificações do Modelo de Dados**;
6. Se o modelo não apresenta os resultados desejados, avalie a possibilidade de retorno a uma Atividade anterior. Avalie a construção de um novo Modelo, inicie mudando a parametrização da ferramenta. Se o resultado ainda não atende, busque outras técnicas de modelagem e reavalie a construção do modelo. Volte à Tarefa 5, Atividade 1 – Selecionar a Técnica de Modelagem.

50 

Atividade 4_ AVALIAR TECNICAMENTE O MODELO DE DADOS

- 1_ Avaliar resultados com respeito a critérios de avaliação do Modelo de dados
- 2_ Interpretar os resultados em termos de negócios – tanto quanto possível nesta fase
- 3_ Obter comentários dos especialistas do domínio sobre os resultados
 Analista de Negócios
- 4_ Verificar a plausibilidade do Modelo
 Analista de Negócios
- 5_ Incrementar as **Especificações do Modelo de Dados**
- 6_ Se o resultados desta avaliação não atenderem os critérios estabelecidos volte à **TAREFA 5, Atividade 1 - SELECIONAR A TÉCNICA DE MODELAGEM** 



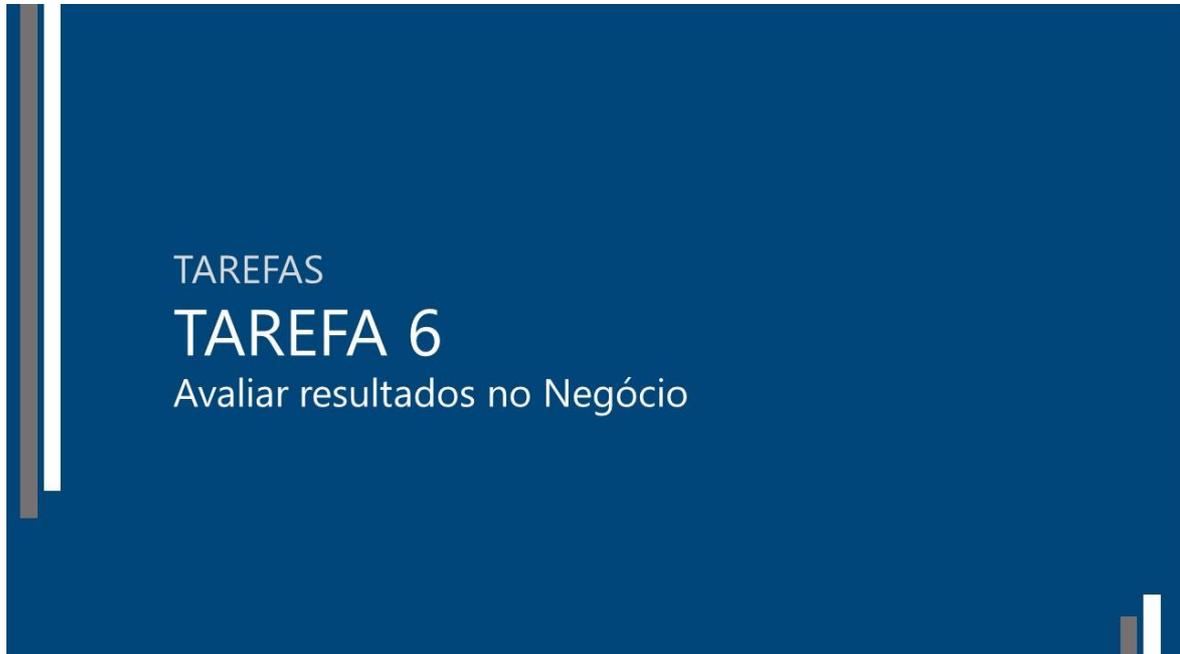
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas
- _ Objetivos do Negócio



- _ Especificações do Modelo de Dados

SLIDE 64

O objetivo deste *slide* é organizar a apresentação.

**SLIDE 65****Tarefa 6 – Avaliar os resultados no Negócio**

Essa etapa avalia o grau em que o Modelo atende aos objetivos do Negócio e procura determinar se existe algum motivo pelo qual esse modelo é deficiente. Os resultados são avaliados de acordo os critérios de avaliação definidos no início do Método. Também é avaliado o processo como um todo, procurando oportunidades de melhoria na execução das suas Tarefas e Atividades.

51 

TAREFA 6

Atividade 1_ Avaliar resultados do Modelo de Dados

Atividade 2_ Revisar as Atividades realizadas

Atividade 3_ Determinar próximos passos

 _ Relação de Recursos Disponíveis
_ Relação de Impactos e Contingências
_ Descrição de Dados
_ Especificações do Modelo de Dados
_ Relação de Ferramentas
_ Objetivos do Negócio

SLIDE 66

Avaliar o resultado do Modelo de Dados

Agora você já tem o Modelo de Dados tecnicamente correto e está na hora de ser avaliado pelo Negócio.

Ações

1. Compreender e interpretar os resultados apresentados pelo Modelo de Dados.
Ação realizada pelo Especialista do domínio;
2. Verificar se o conhecimento obtido é novo e útil para o Negócio;
3. Avaliar os resultados em relação aos critérios de sucesso determinados pelo negócio.

52 

Atividade 1_ AVALIAR O RESULTADO DO MODELO DE DADOS

- 1_ Compreender e interpretar os resultados apresentados pelo Modelo de Dados
 Analista de Negócios
- 2_ Verificar se o conhecimento obtido é novo e útil para o Negócio
 Analista de Negócios
- 3_ Avaliar os resultados em relação aos critérios de sucesso do Negócio
 Analista de Negócios

 _ Objetivos do Negócio
 _ Modelo de Dados

SLIDE 67

Revisar Tarefas

Nesse ponto, o Modelo de Dados já foi avaliado. Conforme o resultado apresentado pelo negócio, satisfatório ou não, torna-se necessário revisar as tarefas e atividades que fazem parte do Método e validar se sua execução foi satisfatória ou se é necessário realizar mudanças.

Ações

1. Fazer uma avaliação geral em função dos resultados obtidos;
2. Avaliar cada Atividade realizada, questionando:
 - Foi necessária?
 - Foi executada de forma otimizada?
 - De que maneiras poderia ser melhorada?
3. Identificar falhas;
4. Identificar possíveis ações alternativas e / ou caminhos inesperados.

53 

Atividade 2_ REVISAR TAREFAS

- 1_ Fazer uma avaliação geral em função dos resultados obtidos  Analista de Negócios
- 2_ Avaliar cada Atividade realizada questionando:  Analista de Negócios
 - _ Foi necessária?
 - _ Foi executada de forma otimizada?
 - _ De que maneiras poderia ser melhorada?
- 3_ Identificar falhas  Analista de Negócios
- 4_ Identificar possíveis ações alternativas e/ou caminhos inesperados  Analista de Negócios



- _ Relação de Recursos Disponíveis
- _ Relação de Impactos e Contingências
- _ Descrição dos Dados
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas
- _ Objetivos do Negócio

SLIDE 68

Determinar os próximos passos

Com base nos resultados da avaliação e na revisão das atividades, é decidida a implantação, ou não, dos resultados obtidos. Também pode ser decidida a execução de novas iterações com mudanças na execução das atividades.

Ações

1. Analisar o potencial de implantação do resultado;
2. Estimar o potencial de melhoria do processo atual;
3. Avaliar opções alternativas de continuidade;
4. De acordo com o resultado da avaliação e se for necessário, decida para qual Tarefa e Atividade é necessário retornar.

54 

Atividade 3_ DETERMINAR PRÓXIMOS PASSOS

- 1_ Analisar o potencial de implantação do resultado  Analista de Negócios
- 2_ Estimar o potencial de melhoria do processo atual  Analista de Negócios
- 3_ Avaliar opções alternativas de continuidade  Analista de Negócios
- 4_ Reiniciar o processo na tarefa de acordo com a avaliação anterior 

TAREFA XX, Atividade XX

 - Relação de Recursos Disponíveis

- _ Relação de Impactos e Contingências
- _ Descrição dos Dados
- _ Especificações do Modelo de Dados
- _ Relação de Ferramentas
- _ Objetivos do Negócio

SLIDE 69

O objetivo deste *slide* é organizar a apresentação.

TAREFAS

TAREFA 7

Implantar a solução

SLIDE 70

Tarefa 7 – Implantar o Modelo

O modelo está pronto e validado. Agora, você precisa incorporar o mesmo aos processos do Negócio e monitorar seu uso.

55 

TAREFA 7

Atividade 1_ Planejar a implantação

Atividade 2_ Planejar monitoramento e manutenção

 _ Objetivos do Negócio

 _ Plano de Implantação

SLIDE 71

Planejar a Implantação

Essa atividade tem como objetivo desenvolver uma estratégia para a Implantação e monitoramento do Modelo de Dados.

Ações

1. Determinar como o conhecimento ou a informação será propagada para os usuários;
2. Decidir como o uso do resultado será monitorado e seus benefícios serão medidos.
3. Desenvolver e avaliar um **Plano de Implantação**;

56 

Atividade 1_ PLANEJAR IMPLANTAÇÃO

- 1_ Determinar como o conhecimento ou a informação será propagada para os usuários
 Analista de Negócios
- 2_ Decidir como o uso do resultado será monitorado e seus benefícios medidos
 Analista de Negócios
- 3_ Desenvolver um **Plano de Implantação**

 _ Objetivos do Negócio

 _ Plano de Implantação

SLIDE 72

Planejar Monitoramento e Manutenção

Monitoramento e manutenção são questões importantes se os resultados da aplicação do Modelo se tornam parte do dia-a-dia do Negócio. Uma preparação cuidadosa de uma estratégia de manutenção ajuda a evitar períodos desnecessariamente longos de uso incorreto dos resultados. Para monitorar a implantação do(s) resultado(s) do modelo, é necessário criar um plano detalhado para monitoramento e manutenção. Esse plano leva em consideração o tipo específico de implantação.

Ações

1. Verificar se há aspectos dinâmicos ou temporais no Negócio que afetem o resultado do Modelo. Ou seja, que coisas podem mudar no ambiente?
2. Definir uma forma (plano) de monitoramento;
3. Determinar quando o resultado ou o Modelo de Dados não deve mais ser usado. Identifique critérios. Exemplos: validade, limite de precisão, novos dados e alteração no domínio da aplicação. O que deve acontecer se o modelo

ou resultado não puder mais ser usado. Exemplo: atualizar modelo ou configurar novo projeto de mineração de dados.

4. Documente estas ações e incremente o **Plano de Implantação**.

57 

Atividade 2_ PLANEJAR MONITORAMENTO E MANUTENÇÃO

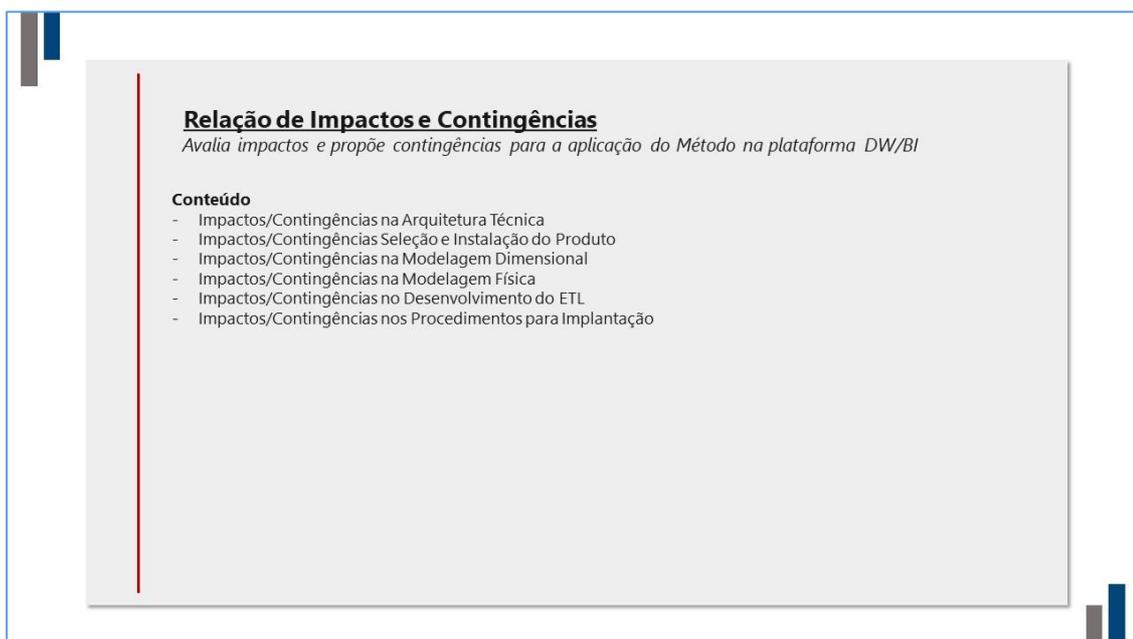
- 1_ Verificar se há aspectos dinâmicos ou temporais no Negócio que afetem o resultado do Modelo
 Analista de Negócios
- 2_ Definir uma forma (plano) de Monitoramento
 Analista de Negócios
- 3_ Determinar quando o resultado ou o Modelo de dados não deve mais ser usado; identificar critérios
 Analista de Negócios
- 4_ Incrementar o **Plano de Implantação**

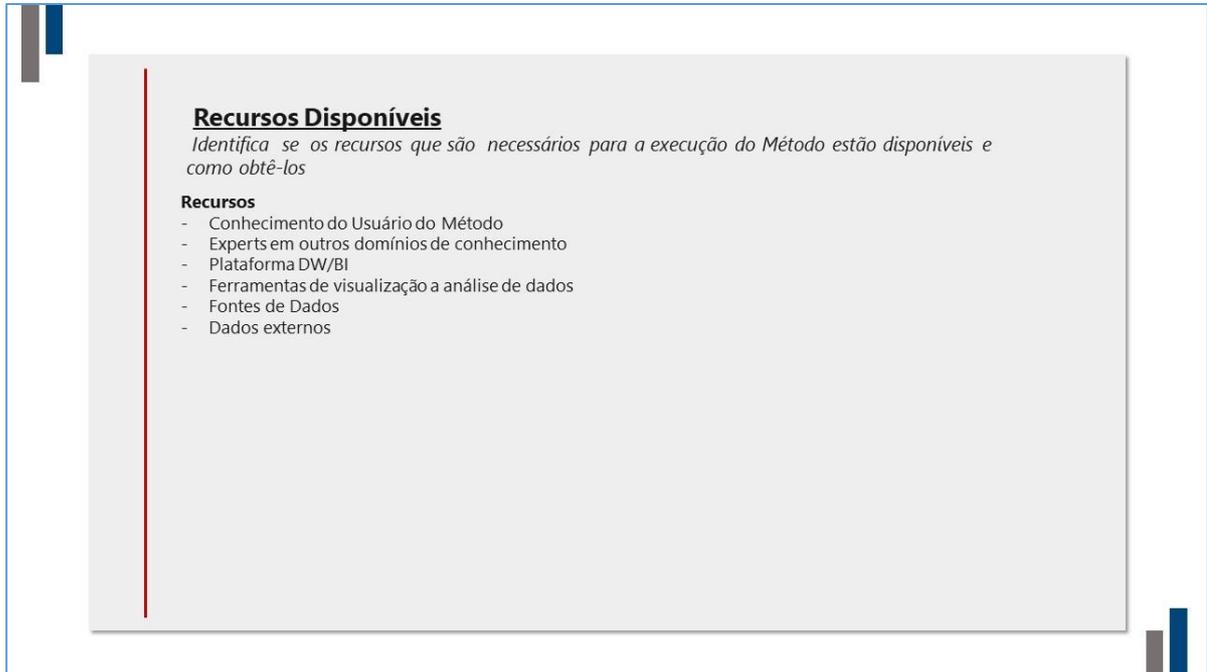

_ Objetivos do Negócio
_ Plano de Implantação


_ Plano de Implantação

SLIDE 73

O objetivo deste *slide* é organizar a apresentação.

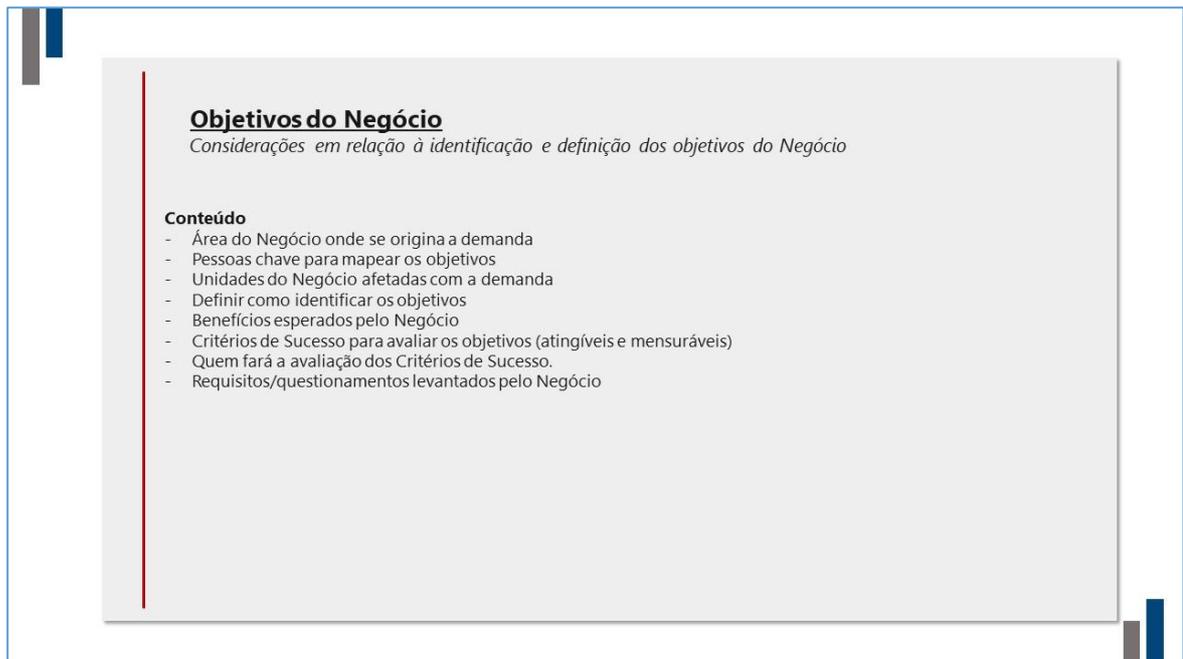
**SLIDE 74**

SLIDE 75The slide content is contained within a light gray rectangular area. On the left side of this area, there is a vertical red line. The text is left-aligned. The slide is framed by a blue border, with decorative vertical bars in gray and blue on the left and right sides.

Recursos Disponíveis
Identifica se os recursos que são necessários para a execução do Método estão disponíveis e como obtê-los

Recursos

- Conhecimento do Usuário do Método
- Experts em outros domínios de conhecimento
- Plataforma DW/BI
- Ferramentas de visualização a análise de dados
- Fontes de Dados
- Dados externos

SLIDE 76The slide content is contained within a light gray rectangular area. On the left side of this area, there is a vertical red line. The text is left-aligned. The slide is framed by a blue border, with decorative vertical bars in gray and blue on the left and right sides.

Objetivos do Negócio
Considerações em relação à identificação e definição dos objetivos do Negócio

Conteúdo

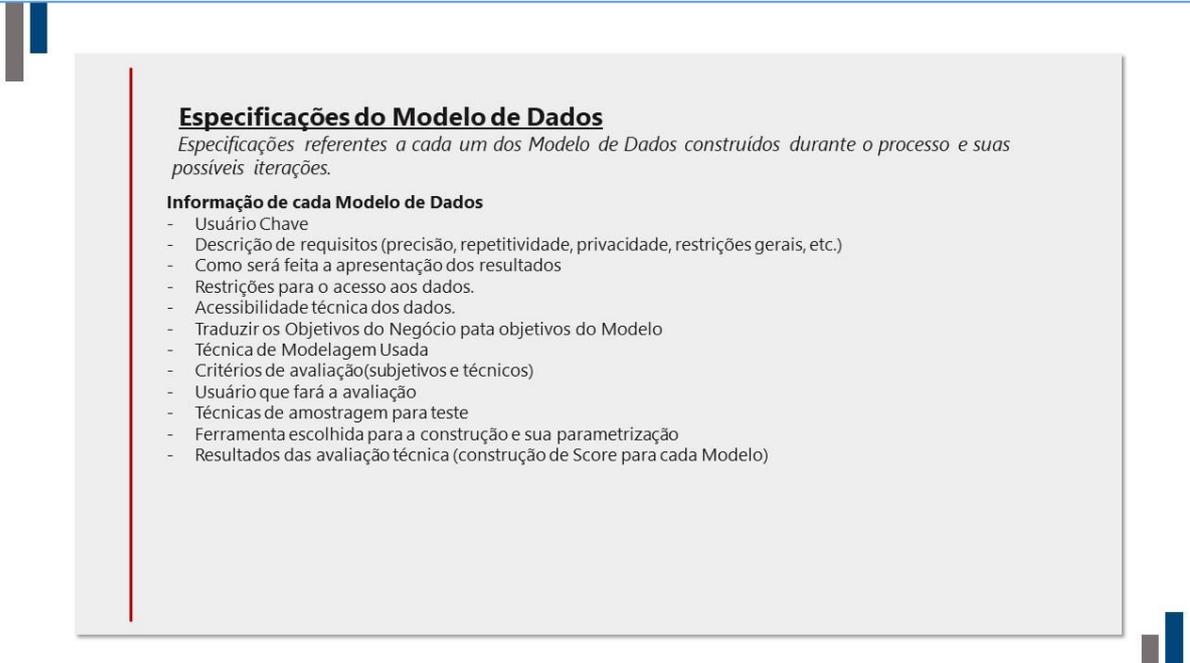
- Área do Negócio onde se origina a demanda
- Pessoas chave para mapear os objetivos
- Unidades do Negócio afetadas com a demanda
- Definir como identificar os objetivos
- Benefícios esperados pelo Negócio
- Critérios de Sucesso para avaliar os objetivos (atingíveis e mensuráveis)
- Quem fará a avaliação dos Critérios de Sucesso.
- Requisitos/questionamentos levantados pelo Negócio

SLIDE 77

Relação de Ferramentas
Ferramentas que podem ser usadas para a construção do Modelo de Dados

Informação de cada Ferramenta

- Nome da Ferramenta
- Funcionalidades
- Técnicas que aplica para Modelagem
- Métricas disponíveis para validar o Modelo
- Tipo de dados que a ferramenta acesso
- Restrições na capacidade de processamento de acordo ao volume de dados
- Requisitos para integração na plataforma DW/BI
- Disponibilidade na organização

SLIDE 78

Especificações do Modelo de Dados
Especificações referentes a cada um dos Modelo de Dados construídos durante o processo e suas possíveis iterações.

Informação de cada Modelo de Dados

- Usuário Chave
- Descrição de requisitos (precisão, repetitividade, privacidade, restrições gerais, etc.)
- Como será feita a apresentação dos resultados
- Restrições para o acesso aos dados.
- Acessibilidade técnica dos dados.
- Traduzir os Objetivos do Negócio para objetivos do Modelo
- Técnica de Modelagem Usada
- Critérios de avaliação (subjetivos e técnicos)
- Usuário que fará a avaliação
- Técnicas de amostragem para teste
- Ferramenta escolhida para a construção e sua parametrização
- Resultados das avaliação técnica (construção de Score para cada Modelo)

SLIDE 79

Descrição dos Dados, 1/2
Descrição dos dados candidatos para a construção do Modelo de Dados.

Geral

- Identificação dos dados necessários (tabelas ou arquivos)
- Validar a disponibilidade dos Dados
- Critérios para seleção de Atributos
- Tempo de uso do histórico
- Identificação dos dados externos necessários

Informação por Atributo

- Tipo de atributo
- Intervalo de valor
- Tabela ou arquivo de origem
- Disponibilidade
- Correlação com outros Atributos
- Significado
- Cálculo de estatísticas básicas
- Relevância

Informação por chave

- Principais relacionamentos
- Sobreposições nas tabelas

**SLIDE 80**

Descrição dos Dados, 2/2

Qualidade dos Dados (para cada dado fazer a validação e documentar as correções)

- Cobertura
- Significado em concordância com os valores contidos
- Atributos ausentes e o significado da sua ausência
- Atributos semelhantes com diferente significado
- Ortografia e formato dos valores
- Ruídos
- Plausibilidade



SLIDE 81**Plano de Implantação**

Implantação acompanhamento e manutenção do Modelo de Dados no contexto do Negócio

- Descrever como propagar o conhecimento entre os usuários
- Descrever como monitorar os resultados da aplicação do Modelo
- Aspectos dinâmicos ou temporais do Negócio que afetam o Modelo
- Identificar (gatilhos) quando os resultados do Modelo não são mais válidos

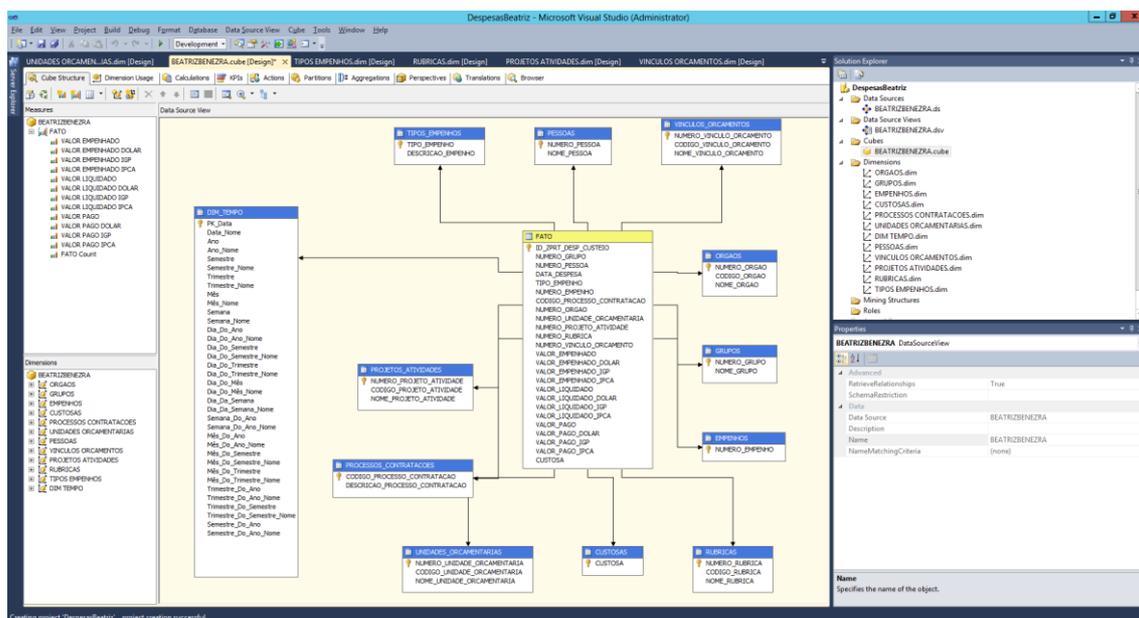
APÊNDICE B

Exercício de aplicação do Método

Definição do Exercício

Visando a transparência organizacional em relação ao uso de recursos, a Prefeitura de Porto Alegre disponibiliza para a cidadão os dados de despesas de custeio. Estes estão disponíveis no portal DATAPOA (www.datapoa.com.br). Existe uma população total de 370.631 registros, organizados em 30 *datasets* referentes às despesas bimestrais de 2010 a 2014.

Os dados foram capturados e trabalhados para criar um modelo OLAP estrela, conforme o diagrama abaixo.



As despesas foram indexadas por Dólares Americanos, IGM-DI/FGV e IPCA/IBGE, para permitir análises com eventual perda do poder aquisitivo do Real.

As consultas disponíveis no portal atualmente são realizadas considerando informações históricas e com o objetivo de verificar algumas hipóteses pré-definidas baseadas em dados históricos. A título de exemplo segue parte destas das consultas disponíveis no *site*:

1. Quais são os valores totais e valores médios, empenhados e efetivamente pagos, por mês e ano e por Órgão, Grupo e Vínculo Orçamento?

2. Quantidade total de despesas por 'Projetos Atividades' no ano de '2014' agrupado por 'Semana'?
3. Qual o valor total liquidado da unidade orçamentária 'SECRETARIA MUNICIPAL DE SAÚDE' no período de 01/06/2013 à 31/05/2014 por rubricas e trimestre?
4. O valor total empenhado agrupado por vínculo orçamentário e por pessoa no ano de 2014?
5. Qual o valor pago mensalmente a cada ano por grupo (pessoal, terceiros, custeio e consumo), em reais e indexado por dólar, IGP e IPCA?
6. Qual o valor pago por rubrica e por tipo de contratação, semanalmente, nos últimos 3 anos?
7. Qual é o Valor Empenhado e Valor Pago para cada Rubrica por Tipo de Processo Contratação a cada mês do ano?
8. Quais os valores pagos no primeiro trimestre do ano de 2012, por projeto ativo?
9. Qual o total do valor pago, por Nome de Projeto Ativo (NOMEPROJATIV) e por Vínculo de Orçamento (NOMEVINCORC), e por Órgão (NOMEORGAO), semestralmente?
10. Quais são as médias de valores empenhados, liquidados e pagos para cada tipo de processo de contratação e por cada processo, em cada semana?
11. Valor liquidado e valor médio empenhado, mensalmente, por grupo no projeto de atividades "carnaval"?
12. Total de valor pago em processos, por órgão e unidade, trimestralmente?
13. Quais são os valores totais e valores médios, empenhados, mensalmente, por Nome do Órgão e Nome do Vínculo Orçamento, para despesas Custosas?
14. Quais são os valores totais empenhados, em Reais e indexados por Dólar Comercial, semanalmente, por Nome da Unidade Orçamentária?
15. Quais os valores médios empenhados, em Reais, trimestralmente, por Nome de Grupo e por Tipo de Empenho, para despesas não Custosas?
16. Quais são os valores totais e valores médios, empenhados, trimestralmente, por Nome do Órgão e Nome do Vínculo Orçamento, para despesas Não Custosas?
17. Quais são os valores totais empenhados, em Reais e indexados por Dólar Comercial, semanalmente, por Nome da Unidade Orçamentária?

18. Quais os valores médios empenhados, em Reais, mensalmente, por Nome de Grupo e por Tipo de Empenho, para despesas Custosas?

Oportunidade para apresentar um comportamento inteligente

As despesas de custeio estão bem exploradas pelo ponto de vista de análise do passado. Contudo, nada se tem para prever o futuro, principalmente para ajudar o governo a planejar as despesas com antecedência e com boa margem de acerto. Também deseja-se conhecer a composição orçamentária de acordo com os principais elementos que definem o orçamento (Unidade Orçamentária, Processo de Contratação, Tipo de Empenho, etc.). Para atingir este objetivo, será necessário aplicar mecanismo de Descoberta de Conhecimento em Banco de Dados, em particular gerar Modelos de Previsão.

É necessário ter à disposição um arsenal maior de tecnologias para permitir induzir modelos de previsão mais sofisticados e que tenham boa margem de acerto no nível de Órgão, Unidade Orçamentária, Vínculo Orçamento, etc.

Ressalvas para a aplicação do Método

As ressalvas para a aplicação do Método forma descritas no Capítulo 5 deste trabalho. Ainda cabe acrescentar:

- A Partir do modelo OLAP foi gerada uma planilha Excel contendo todos os atributos do modelo. Com este objetivo realizou-se um INNER-JOIN das tabelas do modelo estrela.
- Desconhece-se se foi realizado ou não algum outro tratamento de dados antes da geração da estrutura OLAP, como por exemplo o tratamento de dados realizado como parte do processo de ETL.

Execução do Método

A seguir apresenta-se a tabela com as Tarefas/Atividades/Ações que compõem o Método. Para cada Ação descreve-se como foi feita sua aplicação e o resultado obtido. O Método orienta à geração de diferentes relatórios contendo os resultados da sua aplicação. Estes relatórios apresentam-se logo após a tabela de execução do Método.

TAREFA	ATIVIDADE	AÇÃO	APLICAÇÃO	RESULTADO
TAREFA 1	1- Avaliar os Impactos da aplicação do Método.	As ações não estão especificadas já que a atividade como um todo não se aplica.	NÃO SE APLICA. Esta primeira atividade tem como objetivo entender o contexto onde será aplicado o Método e fazer uma avaliação inicial da viabilidade da para a aplicação do mesmo na plataforma DW/BI.	
	2- Conhecer os recursos disponíveis.	1-Identificar o Conhecimento do usuário do Método.	NÃO SE APLICA. Trata-se de um exercício.	
		2-Conhecer a plataforma DW/BI.	NÃO SE APLICA. O ponto de partida para o exercício é uma estrutura OLAP gerada a partir de dados abertos.	
		3-Fazer um inventário das Ferramentas de visualização e análise de dados disponíveis.	Serão usadas ferramentas livres disponíveis para o usuário do Método e que são do seu conhecimento: WEKA e ORANGE.	Incluir a informação na RELAÇÃO DE RECURSOS DISPONÍVEIS
		4- Avaliar a possibilidade de adquirir novas ferramentas.	NÃO SE APLICA. Apenas serão usadas as ferramentas livres disponíveis.	
		5- Fazer Inventário dos dados disponíveis.	Trata-se de dados abertos disponibilizados pela Prefeitura de Porto Alegre.	Incluir a informação na RELAÇÃO DE RECURSOS DISPONÍVEIS
		6- Verificar a possibilidade de poder obter novos dados de fontes externas.	Para a execução deste exercício apenas serão usados os dados fornecidos.	
		7- Gerar uma relação de Recursos Disponíveis.	Gerar o relatório com as informações coletadas durante a execução da Atividade.	Disponibilizar a RELAÇÃO DE RECURSOS DISPONÍVEIS.
	3- Identificar os objetivos do Negócio	1- Identificar a área do Negócio.	NÃO SE APLICA. Trata-se de um exercício.	
		2- Identificar as pessoas chave para mapear os objetivos.	NÃO SE APLICA. Trata-se de um exercício.	
		3- Identificar as áreas do Negócio que são afetadas com os objetivos do KDD.	NÃO SE APLICA. Trata-se de um exercício.	
		4- Definir como identificar os objetivos.	NÃO SE APLICA. Trata-se de um exercício.	
		5- Especificar os benefícios esperados pelo Negócio.	Esta ação depende do Negócio. Porém serão definidos objetivos e benefício necessários para poder executar o exercício.	Incluir informações no relatório de OBJETIVOS DO NEGÓCIO.
		6- Documentar os Objetivos do Negócio.	Gerar relatório de OBJETIVOS DO NEGÓCIO com as informações coletadas durante a execução da Atividade.	Disponibilizar o relatório OBJETIVOS DO NEGÓCIO.
	4- Definir os Critérios de sucesso para o Negócio	1- Especificar os Critérios de Sucesso para o Negócio (atingíveis e mensuráveis).	NÃO SE APLICA. Trata-se de um exercício.	
		2- Indicar quem avalia os Critérios de Sucesso.	NÃO SE APLICA. Trata-se de um exercício.	
		3- Incrementar a documentação com os Objetivos do Negócio.	Incrementar relatório de OBJETIVOS DO NEGÓCIO com as informações coletadas durante a execução da Atividade.	

TAREFA 2	1- Identificar os Requisitos para a construção do Modelo de Dados	1- Identificar o Usuário alvo.	NÃO SE APLICA. Trata-se de um exercício, não existe um usuário alvo presente.	
		2- Captar os requisitos do Modelo (Precisão, Capacidade de Manutenção, Repetibilidade).	Requisitos avaliados: Capacidade de Manutenção, Repetibilidade.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
		3- Captar outros requisitos adicionais (Segurança, Restrições Legais, Privacidade).	Requisitos avaliados: Segurança, Restrições Legais e Privacidade.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
		4- Gerar uma relação de Especificações do Modelo de Dados.	Gerar relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.	Disponibilizar a relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
	2- Avaliar as premissas para a construção do Modelo de Dados	1- Esclarecer todas as suposições vinculadas ao conteúdo do conjunto de dados.	NÃO SE APLICA. Esta atividade pretende esclarecer requisitos não explicitados e que podem estar sendo considerados pelo usuário.	
		2- Esclarecer todas as suposições em relação ao Modelo de Dados.	NÃO SE APLICA. Esta atividade pretende esclarecer requisitos não explicitados e que podem estar sendo considerados pelo usuário.	
		3- Esclarecer suposições e restrições em relação a fatores externos.	NÃO SE APLICA. Esta atividade pretende esclarecer requisitos, premissas ou restrições vinculadas ao contexto de aplicação do Modelo.	
		4- Identificar como os resultados da aplicação do Modelo serão apresentados ao solicitante.	Serão criadas visualizações usando diferentes tipos de gráficos.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
		5- Incrementar a Relação de Especificações do Modelo de Dados.	Incrementar o relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.	Disponibilizar a relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
	3- Avaliar as restrições para a construção do Modelo de Dados.	1- Verificar as restrições gerais.	Não existe nenhum tipo de restrição (legal, recursos, cronograma, etc.) para a construção do Modelo.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
		2- Verificar os direitos de acesso às fontes de dados.	Trata-se de dados abertos.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
		3- Verificar acessibilidade técnica dos dados.	Conta-se com uma tabela Excel onde foi realizado o INNER_JOIN das tabelas que compõem o Modelo OLAP apresentado. Esta tabela contém todos os registros dos anos especificados (2010 – 2014).	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
		4- Verificar as restrições orçamentárias.	NÃO SE APLICA. Trata-se de um exercício.	
		5- Incrementar a Relação de Especificações do Modelo de Dados.	Incrementar o relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.	Disponibilizar a relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
	4- Avaliar a viabilidade para a construção do Modelo de Dados	1- Avaliar a viabilidade de trabalhar com as restrições impostas.	Não foram encontradas restrições.	
		2- Avaliar se efetivamente as premissas são válidas.	Todas as premissas levantadas até o momento são válidas. Constam na definição do exercício.	
		3- Avaliar a possibilidade de contornar as restrições impostas ou as premissas não válidas.	Não existem restrições nem premissas não válidas	
		4- Se a construção do Modelo de Dados não for viável, volte à TAREFA 1/ATIVIDADE 3 - Identificar os objetivos do Negócio.	Não será necessário iterar. A construção do Modelo de Dados é viável.	Continua-se com a execução do Método sem fazer iterações.
	5- Determinar os Objetivos da aplicação do Modelo de Dados	1- Traduzir os objetivos do Negócio para os objetivos do Modelo de Dados.	Aplica a Análise Preditiva para poder capturar tendências a partir de dados históricos.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
		2- Especificar o tipo de técnica de modelagem de dados que será usada na construção do Modelo de Dados.	Para fazer a classificação aplica-se algoritmos que implementam a técnica de Árvores.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS.
3- Incrementar a Relação de Especificações do Modelo de Dados.		Incrementar o relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.	Disponibilizar o relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.	

TAREFA 2	6- Identificar os Critérios de Sucesso da aplicação do Modelo de Dados	1- Especificar os critérios técnicos de avaliação do Modelo de Dados	É usada a Matriz de Confusão associada à Classe Objetivo e também outras variáveis fornecidas pela ferramenta que avaliar o desempenho da árvore na classificação.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		2- Definir Benchmarks para os critérios de avaliação.	A geração de diferentes Modelos de Dados a partir da alteração dos parâmetros para a formação das árvores permitirão estabelecer uma comparação entre os modelos para posteriormente escolher o mais adequado desde o ponto de vista técnico.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		3- Identificar possíveis critérios de avaliação subjetivos.	NÃO SE APLICA. Esta questão está associada ao Negócio.	
		4- Definir as pessoas que farão a avaliação.	NÃO SE APLICA. Esta questão está associada ao Negócio.	
		5- Incrementar a relação de Especificações do Modelo de Dados.	Incrementar o relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
	7- Fazer uma avaliação inicial das ferramentas para construir o Modelo de Dados	1- Identificar as ferramentas de acordo com as técnicas de modelagem que serão usadas.	Tanto WEKA quanto ORANGE permitem aplicar as técnicas de modelagem selecionadas.	Incluir informações no RELATÓRIO DE FERRAMENTAS
		2- Considerar as ferramentas que já estão disponíveis na sua plataforma DW/BI.	NÃO SE APLICA. O exercício está sendo realizado fora de uma plataforma DW/BI.	
		3- Escolher ferramentas e técnicas potenciais.	WEKA e ORANGE com técnicas de classificação e predição.	Incluir as informações no RELATÓRIO DE FERRAMENTAS
		4- Validar os requisitos de HW das ferramentas e ver se é possível colocá-las na plataforma DW/BI.	As ferramentas escolhidas podem ser executadas no Computador Pessoal onde está sendo realizado o exercício.	
		5- Crie uma Relação de Ferramentas contendo as ferramentas potenciais.	Criar o relatório com a RELAÇÃO DE FERRAMENTAS.	Disponibilizar o relatório RELAÇÃO DE FERRAMENTAS.
TAREFA 3	1- Coletar os Dados Iniciais	1- Planejar quais informações são necessárias para elaborar o Modelo de Dados.	Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS.	Incluir as informações no relatório de DESCRIÇÃO DE DADOS.
		2- Verificar se todas as informações necessárias estão disponíveis.	NÃO SE APLICA. As informações disponíveis são aquelas que fazem parte do Modelo OLAP. Deve-se trabalhar com esta restrição.	
		3- Especificar os critérios de seleção dos atributos.	Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		4- Selecionar tabelas e arquivos de interesse.	NÃO SE APLICA. A única tabela disponível é a gerada a partir do Modelo OLAP. Deve-se trabalhar com esta restrição.	
		5- Selecionar dados dentro de uma tabela ou arquivo.	Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		6- Considerar quanto tempo um histórico deve ser usado.	NÃO SE APLICA. Esta questão já está predefinida no exercício. Deve-se trabalhar com esta restrição.	
		7- Validar a necessidade de obtenção de novos dados de fontes externas.	NÃO SE APLICA. Não se considera necessário incluir mais dados externos para validar o exercício.	
		8- Gere um documento de Descrição de Dados.	Criar o relatório DESCRIÇÃO DE DADOS.	Disponibilizar o relatório de DESCRIÇÃO DE DADOS.
	2- Entender os atributos	1- Verificar acessibilidade e disponibilidade de atributos.	Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		2- Verificar o tipo de atributos.	Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		3- Verificar Intervalos de valores dos atributos	Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS	Incrementar o relatório de DESCRIÇÃO DE DADOS
		4- Analisar correlações dos atributos.	Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		5- Entender o significado de cada atributo e o valor do atributo em termos do Negócio.	NÃO SE APLICA. Esta atividade depende da interação com o Negócio.	
		6- Para cada atributo calcular estatísticas básicas.	A ferramenta apresenta uma funcionalidade que faz avaliação estatística dos atributos apresentando: gráfico de dispersão(histograma), centro, % dispersão, valor mínimo, valor máximo e % de valores faltantes. Informações disponíveis na TABELA de ATRIBUTOS que faz parte do relatório de DESCRIÇÃO DE DADOS.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
7- Entrevistar especialistas do domínio para determinar a relevância do atributo.		NÃO SE APLICA. Esta atividade depende da interação com o Negócio.		
8- Incrementar o documento de Descrição de Dados.		Incrementar o relatório de DESCRIÇÃO DE DADOS.	Disponibilizar o relatório de DESCRIÇÃO DE DADOS.	

TAREFA 3	3- Verifique as Chaves	1- Analisar os principais relacionamentos.	NÃO SE APLICA. Resolvidos pelo Modelo OLAP. Não se tem conhecimento do procedimento aplicado.	
		2- Verificar a quantidade de sobreposições dos principais valores de atributos nas tabelas.	NÃO SE APLICA. Resolvidos pelo Modelo OLAP. Não se tem conhecimento do procedimento aplicado.	
		3- Incrementar a Descrição de Dados.	Não existem novas informações a acrescentar.	
		4- Revisar os objetivos para a construção do Modelo de Dados.	Os dados atendam a os objetivos.	
		5- Se os objetivos do Modelo de Dados não forem viáveis volte à TAREFA2/ATIVIDADE 1 - Identificar os Requisitos para a Construção do Modelo de Dados	Não será necessário iterar.	
	4- Avaliar a qualidade dos Dados	1- Considerar a revisão de chaves e atributos.	NÃO SE APLICA. Resolvidos pelo Modelo OLAP. Não se tem conhecimento do procedimento aplicado.	
		2- Verificar a cobertura.	A análise estatística realizada anteriormente apresentou dados de qualidade. Provavelmente os dados foram tratados na formação do Modelo OLAP.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		3- Verificar se o significado dos atributos e valores contidos se encaixam.	A análise estatística realizada anteriormente apoiada na análise dos dados indicam que os valores dos atributos estão adequadas ao conteúdo previsto. Uma análise mais profunda deveria ser realizada junto ao Negócio.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		4- Identificar atributos ausentes e campos em branco.	A análise dos dados realizada anteriormente apresentou dados de qualidade, sem dados ausentes. Provavelmente os dados foram tratados na formação do Modelo OLAP.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		5- Estabelecer o significado de dados ausentes.	De acordo com o resultado da atividade anterior, não é necessário executar esta atividade.	
		6- Verificar se há atributos com valores diferentes e com significados semelhantes.	A análise dos dados realizada anteriormente mostrou atributos correlacionados. Isto permite escolher apenas um deles para elaboração do Modelo. Análises mais profundas em relação ao significado precisariam do apoio do Negócio.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		7- Verificar a ortografia e o formato dos valores.	NÃO SE APLICA. Esta tipo de verificação precisa de apoio do Negócio.	
		8- Verificar se há ruídos.	A Análise Estatística realizada anteriormente apresentou dados de qualidade. Provavelmente os dados foram tratados na formação do Modelo OLAP.	Incrementar o relatório de DESCRIÇÃO DE DADOS.
		9- Verificar a plausibilidade dos valores.	Os valores parecem plausíveis, estão dentro de intervalos válidos ou de acordo com a definição do atributo. Uma análise mais profunda deve ser feita com o apoio do Negócio.	Incrementar o relatório de DESCRIÇÃO DE DADOS
		10- Incrementar o documento Descrição de Dados.	Incrementar o relatório DESCRIÇÃO DE DADOS.	Disponibilizar relatório DESCRIÇÃO DE DADOS. .
	5- Reavaliar o Modelo e formatar suposições para análise futura	1- Formular questões para avaliar o Modelo de Dados.	Refinar informações nas ESPECIFICAÇÕES DO MODELO DE DADOS a partir do conhecimento adquirido sobre os dados. Esta atividade poderia ser executada em forma mais assertiva com o apoio do Negócio.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		2- Transformar as questões em possíveis resultados.	Refinar informações nas ESPECIFICAÇÕES DO MODELO DE DADOS a partir do conhecimento adquirido sobre os dados.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		3- Realizar uma análise básica para avaliar estes resultados.	Refinar informações nas ESPECIFICAÇÕES DO MODELO DE DADOS a partir do conhecimento adquirido sobre os dados. Esta atividade poderia ser executada em forma mais assertiva com o apoio do Negócio.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		4- Esclarecer ou tornar mais precisos os objetivos Modelo de Dados seus critérios de sucesso.	Refinar informações nas ESPECIFICAÇÕES DO MODELO DE DADOS a partir do conhecimento adquirido sobre os dados.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		5- Atualizar a Especificação do Modelo de Dados.	Atualizar relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.	Disponibilizar o relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.
		6- Se os resultados desta análise mostrar a necessidade de fazer mudanças no Modelo de Dados e não forem viáveis, volte à TAREFA 2/ATIVIDADE 1 - Identificar os requisitos para a construção do Modelo de Dados.	O Modelo de Dados definido se apresenta satisfatório para atendendo a demanda do Negócio. Não será necessário iterar e reexecutar atividades vinculadas à geração formação do Modelo de Dados.	Continuar com a execução do Método sem fazer iterações.

TAREFA 4	1- Selecionar os Dados	1- Coletar dados de diferentes fontes.	NÃO SE APLICA. Os dados já estão disponíveis na Tabela Excel grada a partir do Modelo OLAP.	
		2- Considerar a seleção de subconjuntos de Dados diferentes.	NÃO SE APLICA. Os dados já estão disponíveis na Tabela Excel grada a partir do Modelo OLAP.	
		3- Considerar o uso de técnicas de amostragem.	O volume de dados não justifica a execução desta atividade.	
		4- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS.	Não existem informações a acrescentar.	
	2- Limpar Dados	1- Considerar como lidar com qualquer tipo de ruído observado.	Não foram observados Ruídos. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
		2- Corrigir, remover ou ignorar ruído.	Não foram observados Ruídos. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
		3- Decidir como lidar com valores especiais e seus significados.	Não foram observados valores especiais. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
		4- Verificar mecanismos de construção de dados disponíveis com a lista de ferramentas identificadas.	Não será necessário construir dados ausentes. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
		5- Decidir se é melhor realizar a construção dentro da ferramenta ou fora dela.	Não será necessário construir dados ausentes. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
		6- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS.	Não existem informações a acrescentar.	
	3- Construir Dados	1- Verificar os mecanismos de construção disponíveis com a Lista de Ferramentas identificadas.	Não será necessário construir dados novos. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
		2- Decidir se é melhor realizar a construção dentro ou fora da ferramenta.	Não será necessário construir dados novos. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
		3- Incrementar o documento ESPECIFICAÇÕES DO MODELO DE DADOS	Não existem informações a acrescentar.	
	4- Construir atributos e atributos derivados	1- Considerar os seguintes critérios para a construção de atributos. (O conhecimento de fundo nos convence de que algum fato é importante e deve ser representado, embora a ao tenhamos um atributo atual para representá-lo/A algoritmo de modelagem em uso lida apenas com certos tipos de dados/A avaliação do resultado esperado do Modelo sugere que certos fatos não estão sendo cobertos/A adicionar novas informações sobre a importância relevante dos atributos, adicionando novos atributos/Construir ou imputar atributos ausentes/Adicionar novos atributos aos dados acessados/transformar atributos.	Não será necessário construir atributos. Provavelmente este procedimento já foi realizado na construção do Modelo OLAP. Porém, este procedimento e os resultados da sua aplicação é desconhecido.	
	5- Integrar Dados	1- Verificar se as ferramentas de integração são capazes de integrar as fontes de entrada conforme necessária.	Não será necessário integrar fontes de entrada. Os dados já estão em uma tabela Excel.	
		2- Integrar fontes e armazenar resultados.	Não será necessário integrar fontes de entrada. Os dados já estão em uma tabela Excel.	
		3- Reconsiderar Critérios de Seleção de Dados.	Os dados estão adequados.	
		4- Se os resultados desta análise mostrar a necessidade de fazer mudanças no Modelo de Dados e estas não forem viáveis voltar a TAREFA 4/ATIVIDADE 1 - Selecionar Dados.	Não existe a necessidade de fazer mudanças nos dados selecionados. Portanto, não será necessários fazer uma iteração na execução do Método.	Continua com a execução do Método sem fazer iterações.

TAREFA 5	1- Selecionar a Técnica de Modelagem	1- Considerar todo o conhecimento já adquirido sobre os dados.	Usar como referência os relatórios gerados.	
		2- Considerar os resultados esperados.	Usar como referência os relatórios gerados.	
		3- Considerar a Ferramenta.	Usar como referência os relatórios gerados e concluir qual das ferramentas disponíveis é mais adequada.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		4- Selecionar a técnica de modelagem.	Mantém-se a premissa inicial de usar Árvores de Classificação.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		5- Incrementar as Especificações do Modelo de Dados.	Acrescentar as informações identificadas na execução desta atividade no relatório de ESPECIFICAÇÕES DO MODELO DE DADOS .	Disponibilizar o relatório ESPECIFICAÇÕES DO MODELO DE DADOS.
	2- Definir o teste do Modelo de Dados	1- Verificar diferentes testes para cada uma das metas definidas para o Modelo de Dados.	Serão realizados diferentes testes sobre as variáveis técnicas que a ferramenta permite usar para validar a precisão da classificação. Dois Modelos serão criados com diferentes variáveis alvo para fazer esta validação.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		2- Definir os passos necessários.	A ferramenta permite fazer esta validação com a seleção de alguns parâmetros.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		3- Preparar os dados necessários para o teste.	Já que o volume de dados não é muito grande a validação pode ser realizada diretamente sobre todo o volume de dados.	
		4- Selecionar a técnica de modelagem.	Árvores de Classificação.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		5- Incrementar as Especificações do Modelo de Dados.	Acrescentar as informações identificadas na execução desta atividade no relatório de ESPECIFICAÇÕES DO MODELO DE DADOS .	Disponibilizar o relatório ESPECIFICAÇÕES DO MODELO DE DADOS.
	3- Construir o Modelo	1- Definir parâmetros iniciais da ferramenta.	Foram definidos os parâmetros que acompanham o algoritmo de acordo com a ferramenta e capacidade de processamento do equipamento usado.	Ferramenta parametrizada.
		2- Executar a técnica selecionada no conjunto de dados de entrada para produzir o Modelo de Dados.	Técnica Executada.	Modelo construído.
		3- Avaliar o resultado.	Aplica técnicas de visualização, matriz de confusão e parâmetros fornecidos pela ferramenta	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		4- Interpretar o resultado do Modelo.	Aplica técnicas de visualização.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		5- Se os resultados não são satisfatórios construa outro Modelos e compare os resultados. Com este objetivo volte a TAREFA 5/Atividade 3 - Construir o Modelo.	Foi mudada a a variável alvo e o Método foi reiniciado na Tarefa/Atividade indicada.	Deve ser documentada esta mudança no relatório de ESPECIFICAÇÕES DO MODELO DE DADOS.
	4- Avaliar tecnicamente o Modelo de Dados	1- Avaliar os resultados em relação aos critérios de avaliação do Modelo de Dados.	Usadas as métricas fornecidas pela ferramenta: <i>preciso e recall</i> . Também foi gerada a Matriz de Confusão.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		2- Criar ranking de resultados com respeito a critérios de sucesso e avaliação.	Foram gerados dois Modelos diferentes mudando a variável alvo. Estes podem ser avaliados a partir das métricas técnicas já definidas.	Incluir informações na relação de ESPECIFICAÇÕES DO MODELO DE DADOS
		3- Interpretar os resultados em termos do Negócio - tanto quanto possível nesta fase.	Visualizações foram geradas com este objetivo.	As visualizações foram incorporadas ao relatório de ESPECIFICAÇÕES DO MODELO DE DADOS
		4- Obter comentários dos especialistas do domínio sobre estes resultados.	NÃO SE APLICA. Os especialistas não estão presentes	
		5- Verificar a plausibilidade do Modelo.	O Modelo apresenta dados coerentes, porém é necessário a avaliação dos padrões produzidos pelo Negócio	
		6- Se os resultados desta avaliação não atenderem os critérios estabelecidos volte à TAREFA 5- Criar Modelo de Dados.	Foram criados dos Modelos diferentes mediante a mudança da sua estrutura (atributo objetivo) com o intuito de validar o Método e ter elementos de comparação para avaliar o modelo mais adequada a partir dos mecanismos técnicos de avaliação.	Não é necessários reiniciar a execução do Método.

TAREFA 6	1- Avaliar o resultado do Modelo de Dados	1- Compreender e interpretar os resultados apresentados pelo Modelo de Dados.	Visualizações foram geradas com este objetivo. Porém faltaria a presença do Negócio para avaliar os Modelos gerados.	As visualizações foram incorporadas ao relatório de ESPECIFICAÇÕES DO MODELO DE DADOS
		2- Verificar se conhecimento contido e novo e útil para o Negócio.	NÃO SE APLICA. Apenas o Negócio tem esta visão.	
		3- Avaliar os resultados em relação aos critérios de sucesso do Negócio.	NÃO SE APLICA. Apenas o Negócio tem esta visão.	
	2- Revisar Tarefas	1- Fazer uma avaliação geral em função dos resultados obtidos.	Esta avaliação deveria ser feita com todos os envolvidos. De qualquer forma o autor documenta observações na Seção 5 deste trabalho como melhorias identificadas na execução do Método.	Documentado na Seção 5 deste trabalho.
		2- Avaliar cada Atividade realizada questionando: Foi necessária?/Foi executada de forma otimizada?/De que maneiras poderia ser melhorada?	Esta avaliação deveria ser feita com todos os envolvidos. De qualquer forma o autor documenta observações na Seção 6 deste trabalho como melhorias na identificadas na execução do Método.	Documentado na Seção 5 deste trabalho.
		3- Identificar falhas.	Esta avaliação deveria ser feita com todos os envolvidos. De qualquer forma o autor documenta observações na Seção 5 deste trabalho como melhorias na identificadas na execução do Método.	Documentado na Seção 5 deste trabalho.
		4- identificar possíveis ações alternativas e/ou caminhos inesperados.	Esta avaliação deveria ser feita com todos os envolvidos. De qualquer forma o autor documenta observações na Seção 5 deste trabalho como melhorias na identificadas na execução do Método.	Documentado na Seção 6 deste trabalho.
	3- Determinar próximos passos	1- Analisar o potencial de implantação do resultado.	NÃO SE APLICA. Deve ser definido junto ao Negócio.	
		2- Estimar o potencial de melhoria do processo atual.	NÃO SE APLICA. Deve ser definido junto a todos os envolvidos.	
		3- Avaliar opções alternativas de continuidade.	NÃO SE APLICA. Deve ser definido junto a todos os envolvidos.	
		4- Reiniciar o processo na tarefa de acordo com a avaliação anterior . TAREFA XX/ Atividade XX.	NÃO SE APLICA. Esta ação é consequência das anteriores que também não se aplicam.	
	TAREFA 7	1- Planejar a implantação	1- Determinar como o conhecimento ou a informação será propagada.	NÃO SE APLICA. Deve ser definido junto ao Negócio.
2- Decidir como o uso do resultado será monitorado e seus benefícios medidos.			NÃO SE APLICA. Deve ser definido junto ao Negócio.	
3- Desenvolver um plano de implantação.			NÃO SE APLICA. Deve ser definido junto ao Negócio.	
2- Planejar Monitoramento e Manutenção		1- Verificar se há aspectos dinâmicos ou temporais no Negócio que afetem o resultado do Modelo .	NÃO SE APLICA. Deve ser definido junto ao Negócio.	
		2- Definir uma forma (plano) de Monitoramento.	NÃO SE APLICA. Deve ser definido junto ao Negócio.	
		3- Determinar quando o resultado ou o Modelo de dados não deve ser mais usado.	NÃO SE APLICA. Deve ser definido junto ao Negócio.	
		4- Incrementar o Plano de Implantação.	NÃO SE APLICA. Deve ser definido junto ao Negócio.	

Tabela contendo dados da aplicação do Método

Relatório: Recursos Disponíveis

Este relatório identifica se os recursos que são necessários para a execução do Método estão disponíveis e como obtê-los.

- 1- Conhecimento do Usuário do Método - NÃO SE APLICA
O executante é o autor do Método.

- 2- Especialistas em outros domínios de conhecimento – NÃO SE APLICA
Não serão envolvidos outros participantes na execução do exercício.

- 3- Plataforma DW/BI – NÃO SE APLICA
Como ponto de partida para a execução do exercício já foi recebida o modelo OLAP.

- 4- Ferramentas de visualização a análise de dados
 - WEKA: é uma ferramenta de código aberto formada por uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Ela contém ferramentas para preparação de dados, classificação, regressão, *clustering*, mineração de regras de associação e visualização.
 - ORANGE: é uma ferramenta de código aberto para mineração de dados. Ela possui uma programação visual *front-end* para análise de dados explorativa e visualização interativa de dados. Também pode ser usado como uma biblioteca Python. Executa a preparação de dados, testes, aplica algoritmos de aprendizagem de máquina supervisionados e não supervisionados como diferentes tipos de classificação e *clustering*, redes neurais, entre outros.

- 5- Fontes de Dados
Dados de despesas de custeio, disponíveis no portal DATAPOA (www.datapoa.com.br), onde existe uma população total de 370.631 registros.
Os dados questão estão disponíveis em 30 datasets, referentes às despesas bimestrais de 2010 a 2014.

Dados externos

Conforme descrito anteriormente, serão usados dados externos abertos, fornecidos pela Prefeitura de Porto Alegre.

Relatório: Objetivos do Negócio

Considerações em relação à identificação e definição dos Objetivos do Negócio

1- Área do Negócio onde se origina a demanda – NÃO SE APLICA.

Trata-se de um exercício de aplicação do Método realizado pelo seu autor.

Não existe interação com o Negócio.

2- Pessoas Chave para mapear os objetivos – NÃO SE APLICA.

Trata-se de um exercício de aplicação do Método realizado pelo autor do mesmo.

Não existe interação com o Negócio.

3- Unidades do Negócio afetadas com a demanda – NÃO SE APLICA

Trata-se de um exercício de aplicação do método realizado pelo autor do Método.

Não existe interação com o Negócio.

4- Objetivos buscados pelo Negócio

Como será o orçamento do próximo ano e qual será sua composição (órgão, unidade orçamentária, etc.)?

5- Benefícios esperados pelo Negócio

Elaborar um orçamento que atenda a todos os órgãos da prefeitura.

6- Critérios de Sucesso para avaliar os objetivos – NÃO SE APLICA

Trata-se de um exercício de aplicação do Método realizado pelo autor do mesmo.

Não existe interação com o Negócio para poder definir os Critérios de Sucesso.

7- Quem fará a avaliação dos Critérios de Sucesso - NÃO SE APLICA.

Trata-se de um exercício de aplicação do método realizado pelo autor do Método. Não existe uma interação com o Negócio que possa ajudar a fazer uma avaliação crítica dos resultados.

8- Requisitos/questionamentos levantados pelo Negócio.

Avaliar diferentes opções para a construção da estrutura orçamentária do próximo ano.

Relatório: Especificações do Modelo de Dados

Especificações referentes a cada um dos Modelo de Dados construídos durante o processo e suas possíveis iterações. Para execução do exercício são construídos dois Modelos diferentes e que poderão ser avaliados pelos resultados técnicos. A avaliação pelo Negócio não é aplicável neste exercício, porém, é componente fundamental na aplicação do Método. Cada Modelo será gerado a partir da escolha de diferentes variáveis alvo. O Modelo1 usa a variável alvo TIPO DE EMPENHO. O Modelo2 usa a variável alvo CÓDIGO DE UNIDADE ORÇAMENTÁRIA.

1- Usuário Chave – NÃO SE APLICA

Trata-se de um exercício de aplicação do método realizado pelo autor do Método. Não existe interação com o Negócio.

2- Descrição de requisitos

Estes requisitos baseiam-se em suposições do autor, a título de exemplo, já que não existe interação com o Negócio junto ao qual deveriam ser definidos.

- Capacidade de Manutenção: não será realizada nenhuma manutenção do Modelo. Na sua aplicação ele fornecerá um insight para a formação da estrutura orçamentária do próximo ano.
- Repetitividade: o Modelo será aplicado apenas uma vez para a geração da estrutura orçamentária.
- Segurança: não existe nenhuma restrição de segurança, acesso não permitido, referente ao Modelo criado nem aos resultados da sua aplicação. Em relação à segurança de integridade das informações referentes ao modelo e da sua aplicação, serão realizadas cópias backup.
- Restrições Legais: não existe nenhum tipo de restrição legal.
- Privacidade: não existe nenhuma restrição de privacidade em relação ao Modelo nem a sua aplicação.

3- Como será feita a apresentação de resultados

Os resultados serão apresentados com padrões de dados gerados a partir dos Modelos e que deverão ser analisados e validados pelo Negócio.

4- Restrições para o acesso aos Dados

Não existe nenhum tipo de restrição para ao acesso aos dados. Os dados são abertos, estão disponíveis e são de fácil acesso. Não existem custos envolvidos.

5- Acessibilidade técnica dos Dados

Conta-se com uma tabela Excel onde foi realizado o INNER_JOIN das tabelas que compõem o Modelo OLAP apresentado. Esta tabela contém todos os registros dos anos especificados (2010 – 2014).

6- Traduzir os Objetivos do Negócio para objetivos do Modelo.

O objetivo do Negócio determina a criação de um modelo preditivo onde seja possível fazer uma avaliação para formação de uma nova estrutura orçamentária. Algoritmos de classificação permitirão avaliar diferentes composições.

7- Ferramenta escolhida.

A ferramenta escolhida para avaliação, pré-processamento e geração do Modelo de Dados é ORANGE. A ferramenta suporta o volume e tipo de dados que serão usados. Possui algoritmos para tratamento de dados, testes, classificação, predição. Além de uma excelente interface gráfica interativa.

8- Técnica de Modelagem Usada

Poderão ser aplicadas alguma das técnicas, ou combinações delas, relacionadas a d

- Predição – analisa dados históricos para fazer previsões.
- Classificação: – é uma função que mapeia os dados em uma ou várias classes conhecidas. O algoritmo de classificação usado é a ARVORE DE CLASSIFICAÇÃO. Este é um algoritmo simples que divide os dados em nós por pureza de classe. É projetado internamente e pode lidar com conjuntos de dados discretos e contínuos.

9- Crterios de avaliação (subjetivos e tcnicos)

Serão usadas as métricas fornecidas pela ferramenta para o algoritmo selecionado. KNN. Estas métricas são *Precision* (fração de instâncias classificadas que são relevantes) e *Recall* (fração de instâncias relevantes que são recuperadas) além da Matriz de Confusão, realizada sobre a variável considerada objetivo.

10- Usuário que fará a avaliação - NÃO SE APLICA

Trata-se de um exercício de aplicação do método realizado pelo autor do Método.

11- Técnicas de amostragem para teste

De acordo com o volume de dados usados para o exercício, não será necessário usar nenhuma técnica de amostragem.

12- Resultados da avaliação técnica

O Score de Modelo e a Matriz de Confusão são critérios técnicos para seleção do Modelo mais eficiente.

MODELO 1: a variável alvo é o TIPO DE EMPENHO.

Precision: 0.818

Recall: 0.825

A seguir a Matriz de Confusão:

	Predicted	1	2	3	5	6	7	8	10	11	12	13	20	Σ
Actual	1	74.6 %	19.6 %	0.2 %	NA	1.6 %	NA	6.7 %	0.0 %	0.2 %	3.4 %	NA	1.1 %	26384
	2	21.3 %	72.8 %	2.5 %	NA	8.2 %	NA	9.1 %	0.0 %	0.0 %	1.0 %	NA	0.5 %	29730
	3	0.0 %	0.0 %	94.5 %	NA	0.0 %	NA	0.0 %	0.0 %	0.0 %	0.0 %	NA	0.0 %	1992
	5	0.1 %	0.1 %	0.0 %	NA	0.0 %	NA	0.0 %	0.0 %	0.0 %	0.0 %	NA	0.3 %	142
	6	0.3 %	1.1 %	0.0 %	NA	72.7 %	NA	0.5 %	0.0 %	0.0 %	0.0 %	NA	6.4 %	3523
	7	0.1 %	0.2 %	0.0 %	NA	0.0 %	NA	0.0 %	0.0 %	0.0 %	0.0 %	NA	0.0 %	70
	8	2.6 %	5.7 %	2.7 %	NA	0.4 %	NA	83.7 %	0.0 %	0.0 %	0.0 %	NA	1.1 %	47727
	10	0.0 %	0.0 %	0.0 %	NA	0.0 %	NA	0.0 %	84.5 %	10.1 %	0.0 %	NA	0.0 %	44282
	11	0.6 %	0.0 %	0.0 %	NA	0.0 %	NA	0.0 %	14.6 %	86.8 %	0.0 %	NA	0.0 %	9941
	12	0.3 %	0.5 %	0.0 %	NA	0.0 %	NA	0.0 %	0.0 %	1.6 %	95.3 %	NA	0.0 %	1498
	13	0.0 %	0.0 %	0.0 %	NA	0.0 %	NA	0.0 %	0.9 %	1.3 %	0.2 %	NA	0.0 %	489
	20	0.0 %	0.0 %	0.0 %	NA	17.2 %	NA	0.0 %	0.0 %	0.0 %	0.0 %	NA	90.7 %	26951
	Σ	22950	26900	2103		1337		54037	52097	2523	1311		29471	192729

Matriz de Confusão do Modelo 1

MODELO 2: a variável alvo é o CÓDIGO DA UNIDADE ORÇAMENTÁRIA

Precision: 0.921

Recall: 0.942

A seguir a Matriz de Confusão (dividida em duas partes):

	2400	2500	2601	2700	2801	2900	3101	3103	4000	5000	6001	7000	7002	7101	Σ
Actual	201	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2962
	203	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	48
	301	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1392
	400	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	3114
	601	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1926
	800	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1628
	900	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1658
	1001	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2348
	1003	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	4642
	1200	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2365
	1301	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2826
	1401	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	5522
	1406	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	339
	1501	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1950
	1502	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	5591
	1601	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1880
	1700	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	687
	1801	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	19914
	1804	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	31282
	1900	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	303
	1901	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1246
	2001	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2888
	2100	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	8221
	2301	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	3244
	2304	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1000
	2400	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1149
	2500	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	758
	2601	0.0 %	0.0 %	99.9 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	1299
	2700	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	907
	2801	0.0 %	0.0 %	0.0 %	0.0 %	69.1 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	547
	2900	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	940
	3101	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	8.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2992
	3103	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	92.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	7345
	4000	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	16112
	5000	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	0.0 %	0.0 %	4449
	6001	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	98.7 %	0.0 %	0.0 %	0.0 %	37779
	7000	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	67.9 %	5.5 %	0.0 %	1112
	7001	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	13.0 %	19.9 %	0.0 %	1087
	7002	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	14.0 %	54.7 %	0.0 %	2718
	7003	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	5.1 %	20.0 %	0.0 %	994
	7101	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	47.0 %	0.0 %	331
	Σ	1149	758	1300	907	792	2354	7983	16112	4449	38269	1262	4649	704	192729

Matriz de Confusão Modelo 2

Relatório: Relação de Ferramentas

Ferramentas que podem ser usadas para a construção do Modelo de Dados

WEKA

Funcionalidades:

Possui uma coleção de algoritmos para mineração de dados e oferece a possibilidade de executar tarefas de pré-processamento (preparação de dados) e visualização.

Técnica para Modelagem:

Weka permite aplicar diferentes técnicas de Modelagem de dados: Classificação, Clusterização e Associação. Para cada uma destas técnicas são oferecidos diversos algoritmos

Métricas disponíveis para validar o Modelo:

Para cada algoritmo apresentado existem métricas para sua avaliação.

Tipo de dados que a ferramenta acessa:

Permite trabalhar com diferentes estruturas de dados. Para este exercício foi criada uma tabela Excel que pode ser suportada pela ferramenta.

Em relação ao tipo de atributos, dependerá do algoritmo escolhido. Permite fazer transformação/tratamento de dados aplicando diferentes técnicas.

ORANGE

Funcionalidades:

Visualização interativa que facilita a Análise da Dados.

Técnica para Modelagem:

Orange permite aplicar diferentes técnicas de Modelagem de dados: Classificação, Clusterização e Associação. Para cada uma destas técnicas são oferecidos diversos algoritmos para os quais podem ser escolhidas diferentes visualizações.

Métricas disponíveis para validar o Modelo:

Para cada algoritmo apresentado existem métricas para sua avaliação.

Tipo de dados que a ferramenta acessa:

Permite trabalhar com diferentes estruturas de dados. Para este exercício foi criada uma tabela Excel que pode ser suportada pela ferramenta.

Em relação ao tipo de atributos, vai depender do algoritmo escolhido.

Relatório: Descrição dos Dados

Descrição dos dados candidatos para a construção do Modelo de Dados.

Identificação dos dados necessários

Será usada a tabela definida a partir do Modelo OLAP apresentado. Esta tabela conta com 371.600 registros. Os dados já foram tratados como parte do pré-processamento para formação do Modelo OLAP. Porém, as ações realizadas para o pré-processamento são desconhecidas.

Validar a disponibilidade dos Dados

Os dados referentes às tabelas do Modelo OLAP estão disponíveis.

Critérios para seleção de atributos

- Como valores apenas será escolhido o valor do empenho em dólar.
- Campos descritivos das tabelas não serão considerados para a formação do Modelo de Dados.
- Será considerada a data de execução orçamentária o ano e o mês. Outras informações sobre as datas não serão consideradas.
- Apenas um dos atributos correlacionados será usado na construção do Modelo de Dados.
- Atributos não relevantes não farão parte do Modelo de Dados.

Tempo de uso histórico

Anos 2010 -2014

Identificação dos dados externos necessários

Os dados que fazem parte do Modelo OLAP são externos. Novo dados não serão considerado.

Informação sobre chaves

As chaves foram tratadas e os relacionamentos gerados para elaboração do Modelo OLAP.

Informação por Atributo

- Tipo de atributo: documentado na Tabela de Atributos.
- Tabela ou arquivo de origem: documentado na Tabela de Atributos
- Disponibilidade: os atributos do Modelo OLAP estão disponíveis.
- Correlação com outros atributos: análise realizada com a ferramenta Orange e se especificada na Tabela de Atributos
- Significado: documentado na Tabela de Atributos

- Relevância: documentado na Tabela de Atributos. Atributos considerados como não relevantes não serão considerados na construção do Modelo de Dados.

- Cálculo de estatísticas básicas: a análise estatística só foi realizada para os atributos relevantes. Como resultado desta análise constatou-se:

- Não existem *outlayers*

- O único atributo zerado/nulo é PESSOA_NUMERO.

Mais resultados sobre a análise estatística estão documentados na Tabela de Atributos.

Qualidade dos Dados

Os dados já passaram por um processo de validação de qualidade (não conhecido) na formação do Modelo OLAP. A partir da análise estatística constatou-se que os dados são de qualidade, apresentando valores plausíveis. Apenas um dos campos apresentou valores faltantes. Foram identificadas correlações entre os atributos e desta forma apenas um destes atributos será usado para a formação do Modelo de Dados.

A seguir a Tabela com a análise realizada sobre os atributos. Com o objetivo de facilitar a sua visualização a Tabela foi dividida verticalmente e duplicada, aparece em ambas partições, as colunas que identificam a tabela e o atributo.

Tabela de Atributos

TABELA	ATRIBUTO	TIPO	VALORES POSSÍVEIS	DESCRIÇÃO	RELAVÂNCIA	ANÁLISE DE CORRELAÇÃO
TIPOS_EMPENHO	TIPO	INT	1-8	Código do tipo de empenho	SIM	
	DESCRIÇÃO	CHAR		Descrição do tipo de empenho	NÃO	
PROJETOS_ATIVIDADES	NUMERO	INT	1-999	Identifica um projeto em particular	NÃO	ATIBUTOS CORRELACIONADOS
	CODIGO	INT	0-9999	Identifica o código da atividade do projeto	SIM	
	DESCRIÇÃO	CHAR		Descrição da atividade do projeto	NÃO	
PESSOA	NUMERO	INT	0-99999	Código da Pessoa (PJ ou PF) titular do contrato	SIM	
	NOME	CHAR		Nome da Pessoa Titular do Contrato	NÃO	
PROCESSOS_CONTRATAÇÃO	CODIGO	CHAR	CC/TP/SP/CL/F P/PR/FP/SF/IX/ CV	Código do tipo de contratação	SIM	
	DESCRIÇÃO	CHAR		Descrição do tipo de contratação	NÃO	
VINCULO_ORÇAMENTARIO	NUMERO	INT	1-999	Indica vinculo com outro Grupo orçamentário	NÃO	ATIBUTOS CORRELACIONADOS
	CODIGO	INT	1- 9999	Indica vinculo com outro Grupo orçamentário	SIM	
	NOME			Descrição do Vinculo ou se e despesa livre	NÃO	
DIM_TEMPO	PK_DATA	dd/mm/aaaa		DATA da movimentação orçamentária	SIM	
	ANO	dd/mm/aaaa		ANO da movimentação orçamentária	NÃO	
	MÊS	dd/mm/aaaa		MÊS da movimentação orçamentária	NÃO	
				Outros campos das tabela não serão considerados		
CUSTOSA	CUSTOSA	CHAR	SIM/NÃO		SIM	
RUBRICA	NUMERO	INT	1-999	Código da rubrica contábil	SIM	
	CODIGO			Código contábil	NÃO	
	NOME	CHAR		Descrição da rubrica contábil	NÃO	
UNIDADE-ORÇAMENTARIA	NUMERO	INT	1- 999	Identifica a Unidade Municipal	NÃO	ATIBUTOS CORRELACIONADOS
	CODIGO	INT	1- 9999	Identifica a Unidade Municipal	SIM	
	NOME	CHAR		Descrição da Unidade Municipal	NÃO	
GRUPO	NUMERO	INT	1-4	Numero identifica grupo de despesas	SIM	
	NOME	CHAR		Descrição do Grupo de Despesas	NÃO	
ORGÃO	NUMERO	INT	1-99	Numero do órgão responsável pelo orçamento	NÃO	ATIBUTOS CORRELACIONADOS
	CODIGO	INT	1-9999		SIM	
	NOME	CHAR		Nome do Órgão	NÃO	
EMPENHO	EMPENHO	INT	1-9999	Numero do empenho, identifica o empenho	SIM	
FACTO	VALOR_EMPENHADO_DOLAR	NUM		Valor do empenho em dólar	SIM	
				Outros campos das tabela não serão considerados		

TABELA	ATRIBUTO	ANÁLISE ESTATÍSTICA						TRATAMENTO DO ATRIBUTO
		CENTRO	% DISPERSÃO	MIN	MAX	# Categorias	% VLR. FALTANTE	
TIPOS_EMPENHO	TIPO	8,31	0,7	1	2000	12	12	Transformado para Categórico
	DESCRIÇÃO							
PROJETOS_ATIVIDADES	NUMERO							Transformado para Categórico
	CODIGO	2868,99	1,92	1	7903		527	Transformado para Categórico
	DESCRIÇÃO							
PESSOA	NUMERO	14595	0,6	2	3890		0	
	NOME							
PROCESSOS_CONTRATAÇÃO	CODIGO							Transformado para Categórico
	DESCRIÇÃO	Campo avaliado a partir do Histograma				17	0	
VINCULO_ORÇAMENTARIO	NUMERO							
	CODIGO	949,94	1,92	100	7903	103	0	
	NOME							
DIM_TEMPO	PK_DATA						0	
	ANO							
	MÊS							
	Outros campos das tabela não serão considerados							
CUSTOSA	CUSTOSA						2	0
RUBRICA	NUMERO	223,15	0,76	1	454	454	0	Transformado para Categórico
	CODIGO							
	NOME							
UNIDADE-ORÇAMENTARIA	NUMERO							Transformado para Categórico
	CODIGO	24,2	0,6	1	63	56	0	
	NOME							
GRUPO	NUMERO	2,7	0,34	1	4	4	0	Transformado para Categórico
	NOME							
ORGÃO	NUMERO							Transformado para Categórico
	CODIGO	2,73	0,34	1	32	31	0	
	NOME							
EMPENHO	EMPENHO	28581,99	1,03	1	8933		0	
FACTO	VALOR_EMPENHADO_DOLAR							
	Outros campos das tabela não serão considerados							

Visualizações dos Modelos de Dados

Observação: a ferramenta usada permite trabalhar com as visualizações em forma iterativa, fazendo diferentes combinações atributos e tipos de exibições, individualizando dados e valores, colocando etiquetas, etc. Todos estes recursos permitem fazer um profundo trabalho de análise sobre as informações apresentadas. A título de exemplo apresentam-se algumas imagens sobre as quais o Negócio pode realizar o trabalho de análise dos resultado.

A imagen a seguir apresenta alguns nós da árvore referente ao Modelo 1. O percurso da raiz até a folha representa uma regra de classificação. A folha individualizada (contorno em negrito) corresponde à seguinte classificação:

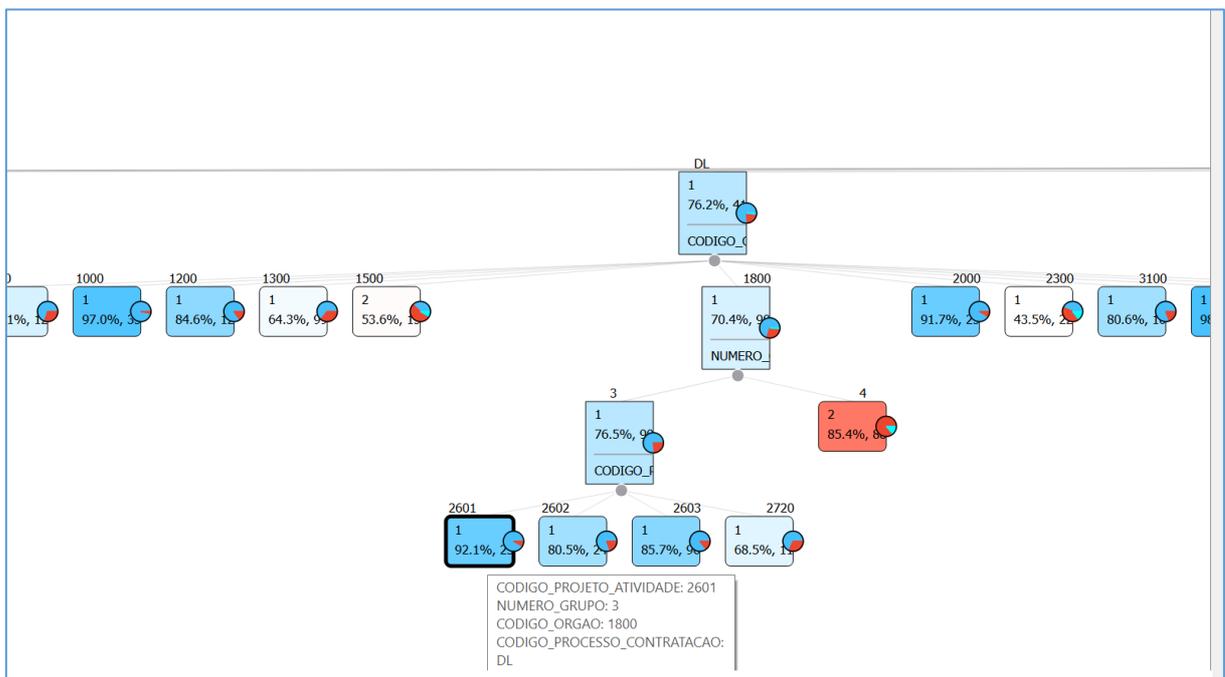
Codigo do Processo de Contratação: DL – Dispasa de Licitação

Numero do Grupo Orçamentário: 3 – Outros Serviços de Terceiros

Código Orgão: 1800 – Administração Geral

Código da Atividade do Projeto: 2601 – Secretaria Municipal da Saúde

Para esta folha a frquência relativa (entropoia) é de 92.1 %



Visão parcial da Árvore de Classificação correspondente ao Modelo 1

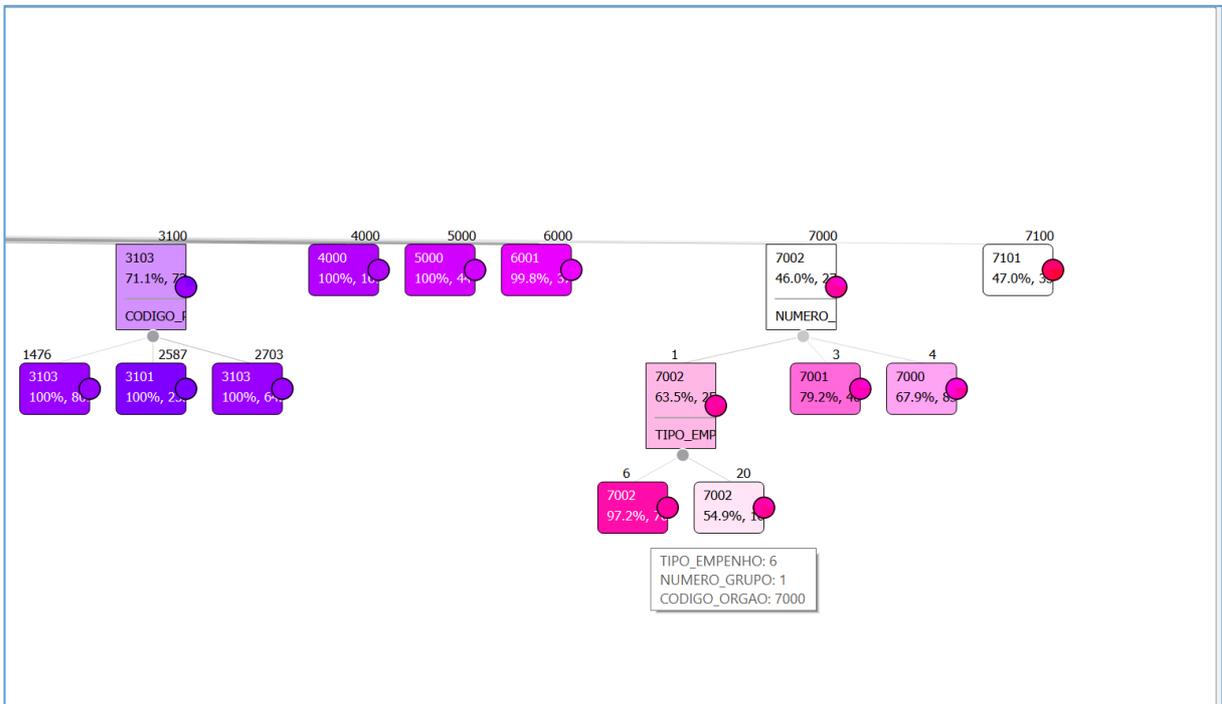
A imagen a seguir apresenta alguns nós da árvore referente ao Modelo 2. O percurso da raiz até a folha representa uma regra de classificação. A folha individualizada (contorno em negrito) corresponde à seguinte classificação:

Tipo de Empenho: 6 – Pessoal

Numero do Grupo Orçamentário: 1 – Pessoal

Código Orgão: 7000 – Previmpa (Previdência Municipal)

Para esta folha a frequência relativa (entropia) é de 97.2 %



Visão parcial da Árvore de Classificação correspondente ao Modelo 1

APÊNDICE C

Questionários para avaliar o Método

QUESTIONÁRIO INICIAL

Nome: _____

Responda ao questionário abaixo, assinalando com "X", de acordo com o seu nível de conhecimento em cada um dos domínios relacionados.

Nenhum	Pouco	Regular	Bom	Muito Bom
1	2	3	4	5

DOMÍNIOS DE CONHECIMENTO						
		1	2	3	4	5
A	Técnicas para Interação com o Negócio					
B	Administração dos dados no Data Warehouse					
C	Arquitetura e desenvolvimento do ETL					
E	Arquitetura e Infraestrutura da plataforma DW/BI					
F	Administração dos Metadados na Plataforma DW/BI					
G	Técnicas e Ferramentas de Mineração de Dados					
H	Desenvolvimento de aplicações BI					
I	Ferramentas BI					
J	Arquitetura de Dados					
K	Preparação e Modelagem de Dados para BI					

QUESTIONÁRIO PARA AVALIAÇÃO DO TUTORIAL

Nome: _____

Empresa: _____

Tempo na Empresa: _____

Marque com "X" uma ou mais opções que melhor representa(m) sua função atual: () Analista de BI, () Analista de Negócios, () Administrador de Dados, () Arquiteto de ETL, () Arquiteto Técnico, () Gestor de Metadados, () Especialista em Mineração.

Tempo na Função: _____

Responda os questionários abaixo de acordo com sua percepção sobre a apresentação do Tutorial. Assinale com "X" de acordo com a seguinte escala.

Discordo Totalmente	Discordo	Não Concordo nem Discordo	Concordo	Concordo Totalmente
1	2	3	4	5

UTILIDADE						
		1	2	3	4	5
A	O Método simplifica muito as tarefas.					
B	O Método direciona claramente as tarefas.					
C	O Método aumenta muito o desempenho.					
D	O Método enedreça claramente seu objetivo.					
E	O Método economiza muito tempo.					
F	O Método permite executar as tarefas com rapidez.					
G	O Método endereça aspectos muito críticos.					
H	O Método é fundamental para conseguir completar as tarefas.					
I	O Método é um grande facilitador.					
J	O Método é extremamente útil.					

FACILIDADE DE USO						
		1	2	3	4	5
A	O Método não gera nenhum tipo de confusão.					
B	O Método não provoca nenhum tipo de erro.					
C	O Método não me decepciona em forma alguma.					
D	O Método não demanda um grande esforço mental.					
E	O Método não é, em forma alguma, rígido ou inflexível.					
F	O Método é muito claro.					
G	O Método não gera situações inesperadas.					
H	O Método não é cansativo em forma alguma.					
I	O Método é de muito fácil entendimento.					
J	O Método é muito fácil de ser aprendido.					
K	O Método me fornece orientações muito úteis.					
L	O Método é, acima de tudo, muito fácil de usar.					



Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br