

DESENVOLVIMENTO DE UMA FERRAMENTA DE AUXÍLIO A TOMADA DE DECISÃO COM BASE NA CRISP-DM.

Rodrigo Rhameses Ribeiro Andrade (IC), Prof. Dr. Alexandre Ferreira de Pinho (PQ)
Universidade Federal de Itajubá.

Palavras-chave: CRISP-DM. Dashboard. Data Mining. ETL.

Introdução

A tomada de decisão é um processo de escolha sobre algo importante especialmente em um grupo de pessoas ou em uma organização. Ela envolve a seleção de um curso de ação entre duas ou mais alternativas possíveis, a fim de se chegar a uma solução para um dado problema. Em simples palavras, a tomada de decisão é o processo de fazer escolhas, identificando uma decisão, reunindo informações e avaliando resoluções alternativas (SILVA, 2009).

A tomada de decisão é parte integrante da gestão moderna. Essencialmente, a tomada de decisão racional ou correta, considerada como função primária de um negócio. Dentro deste contexto, ter conhecimento sobre dados, informações e mineração de dados não só facilita como também diminui a probabilidade de decisões equivocadas, decisões que podem gerar prejuízos e perda de espaço no mercado. O termo dados é apresentado como qualquer elemento identificado em sua forma bruta e que por si só não conduz para a compreensão de um fato ou situação (O'BRIEN, 2002). Já o termo informação é conceituado como uma coleção de fatos organizados e processados de modo que tenham valor adicional (STAIR e REYNOLDS, 2016). A mineração de dados é definida como o processo de descoberta de padrões em dados. O processo deve ser automático ou semiautomático. Os padrões descobertos devem ser significativos na medida em que levam a algum vantagem. Os dados estão sempre presentes em quantidades substanciais (WITTEN et al., 2017).

Dentro do contexto exposto acima, esta pesquisa buscará criar uma ferramenta de auxílio a tomada de decisão com base na modelagem CRISP-DM com o processo ETL e gerar um DASHBOARD com informações que vão dar aos gestores artificiais para tomarem as devidas decisões. Esta pesquisa fez uso da plataforma Excel, da empresa Microsoft.

Esta pesquisa tem o objetivo de usar a modelagem CRISP-DM, e o processo ETL (extração, transformação e carregamento) para criar uma ferramenta de auxílio a tomada de decisão que terá utilidade no dia a dia do negócio. Esta ferramenta buscará ter a capacidade de gerar informações úteis e de grande valia para os gestores e tomadores de decisão.

Como todo e qualquer negocio tem ligação com a tecnologia é importante saber como, onde e de que maneira coletar dados, e posteriormente saber como armazená-los e tratá-los, ou simplesmente, saber utilizar o processo ETL.

Esta pesquisa se faz justa, tendo em vista que, estar conectado com as evoluções tecnológicas e aplicá-las no meio empresarial dá ao negócio meios de sempre evoluir. A pesquisa buscou facilitar a visualização das informações, às organizando em um DASHBOARD interativo, prático e conectado diretamente como a necessidade da empresa que foi usada como exemplo nesta pesquisa.

Lembrando sempre que, um negócio busca sempre eficiência e eficácia, e a ferramenta busca mostrar quais caminhos podem ser tomados para garantir que o negócio alcance o que busca no mercado.

Metodologia

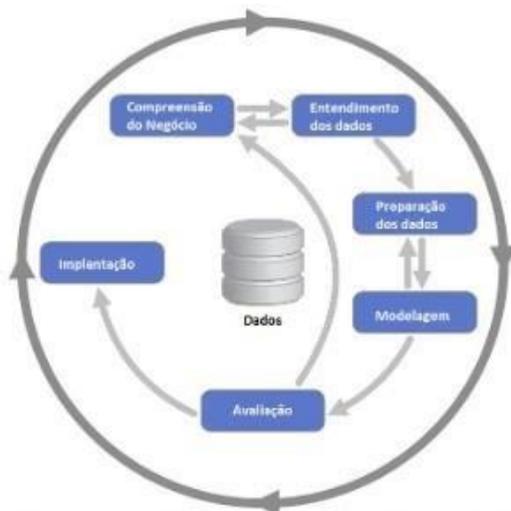
O desenvolvimento da pesquisa, focou em Data Mining, que é parte do Data Science, e que faz uso de estatísticas matemáticas com base em cruzamento de dados, por meio de técnicas de indução para propor hipóteses e solucionar questões empresariais, ou seja, é a mineração de dados que consegue transformar todo o volume de dados obtidos em informações úteis.

A modelagem CRISP-DM, nada mais é que, uma metodologia capaz de transformar dados do negócio em conhecimento e informações de gerenciamento. Essa modelagem existe a mais de vinte anos e teve seu início devido à uma necessidade dos profissionais da área de tecnologia da informação. Quando o assunto foi

transformar dados em informações, os mesmos deixaram a desejar devido ao alto volume de dados. Com isso a CRISP-DM surgiu para atender aos projetos que são desenvolvidos com as análises de um grande volume de dados.

Esta metodologia reúne as melhores práticas para que o Data Mining seja o mais produtivo e eficiente o possível. Analisa dados financeiros, de recursos humanos, de produção e dos clientes, para ter a capacidade de propor modelos de melhoria ou solução de problemas.

A modelagem CRISP-DM é uma metodologia ágil, útil em cenários de incertezas, principalmente na hora de resolver problemas de negócio. Pode ser definida como uma metodologia flexível, capaz de lidar com problemas complexos envolvendo uma grande quantidade de dados. A metodologia CRISP-DM define o ciclo de vida do projeto o dividindo em seis etapas.



Fases do CRISP-DM. Fonte: [Shearer, 2000]

1- Compreensão do Negócio

Primeiro tem-se que entender o que de fato é o problema, buscando todos os detalhes sobre o impacto dele na empresa e quais os objetivos em relação ao trabalho. Com certeza uma etapa importantíssima pois o não entendimento do negócio pode alterar os resultados finais. O CRISP-DM fornece uma visão geral do ciclo de vida da mineração de dados (CHAPMAN, 2000).

2- Entendimento dos Dados

Nesta etapa a organização e documentação de todos os dados disponíveis deve ser minuciosa, é aqui que começa de fato o trabalho de Data Mining, pois a empresa deve ser capaz de identificar quais dados são importantes para a resolução do problema, ou seja, compreender os dados

nós leva a entender que caminhos tomar para tratar os dados e finalmente obter as informações desejadas. Essa etapa é considerada como uma das mais importantes do processo, a partir da qual é desenvolvido um plano preliminar do projeto de mineração para busca dos objetivos (SHEARER, 2000).

3- Preparação dos Dados

Após a identificação dos dados, análise e documentação, pode-se aplicar a técnica de análise nos mesmos. Será nesta etapa que irá ser preparados os Datasets para formatar as questões técnicas da análise, nesta etapa será feita a escolha dos dados e como os mesmos serão cruzados. Essas tarefas incluem a seleção de tabelas, registros e atributos, assim como a transformação e a limpeza dos dados (CHAPMAN et al., 2000).

4- Modelagem

É nesta etapa que será aplicada as técnicas que tem por objetivo identificar o que precisa ser melhorado e ou alterado no negócio, tais técnicas podem ser diferenciadas de acordo com os problemas a serem abordados, esta etapa pode ser executada várias vezes, inclusive, é comum também retornar para a fase de preparação de dados para manipulações exigidas pelo modelo (SHEARER, 2010; IBM, 2016).

5- Avaliação

Esta etapa consiste em realizar teste com o modelo gerado para verificar se os mesmos atendem as necessidades do negócio, também é possível verificar se algum objetivo não foi alcançado com a modelagem proposta, em outras palavras, os dados já foram tratados e transformados em informações que podem ser ou não de utilidade para os tomadores de decisão. Caso não seja satisfatório é possível voltar a etapa inicial. A seta direcionada à etapa inicial indica a possibilidade de retorno em função da avaliação dos resultados não se mostra satisfatória para os objetivos do projeto (CHAPMAN et al., 2000; IBM, 2016).

6- Implementação

Esta é a última etapa é onde se faz a implementação de fato, onde os novos conhecimentos serão aplicados para proporcionar as melhorias no negócio, nesta etapa o conhecimento obtido deve ser organizado de uma forma clara e objetiva. Esta etapa é onde os tomadores de decisão vão de fato utilizar as informações. Dependendo

dos requisitos, a fase de implantação pode ser tão simples como gerar um relatório ou tão complexo como a implementação de um processo de mineração de dados aplicável em toda a empresa (SHEARER, 2000; CHAPMAN et al., 2000).

ETL – EXTRAÇÃO, TRANSFORMAÇÃO E CARREGAMENTO

As decisões gerenciais são tomadas com base nas informações geradas pela modelagem utilizada no negócio. Estas informações são geradas através dos dados armazenados no Dataset. Se estes dados não forem corretamente trabalhados no processo de extração, as informações geradas através deles farão com que decisões sejam tomadas erroneamente, podendo afetar diretamente os negócios. Portanto, os dados devem representar a verdade, a mais pura verdade, nada mais que a verdade (KIMBALL, 1998 apud ABREU, 2007). A maior parte do esforço exigido no desenvolvimento de um Dataset é consumido neste momento e não é incomum que oitenta por cento de todo esforço seja empregado no processo de ETL, (INMON, 1997 apud ABREU, 2007). Extrair, transformar e carregar (ETL) é o processo que as organizações orientadas a dados usam para coletar dados de várias fontes e reuni-los para dar suporte à descoberta, à geração de relatórios, à análise e à tomada de decisões. As origens de dados podem ser muito diversas em tipo, formato, volume e confiabilidade, de modo que os dados precisam ser processados para serem úteis quando reunidos. ETL é um processo dividido em três passos que dão nome ao processo.

1 – Extração

Durante a extração, o ETL identifica os dados e os cópia de suas origens, de forma que possa transportar os dados para o armazenamento de dados de destino. Os dados podem vir de fontes estruturadas e não estruturadas, incluindo documentos, e-mails, aplicações de negócios, bancos de dados, equipamentos, sensores e terceiros.

2 – Transformação

Como os dados extraídos são brutos em sua forma original, eles precisam ser mapeados e transformados para prepará-los para o armazenamento de dados eventual. No processo de transformação, o ETL válida, autentica, agrega os dados de formas que tornam os dados resultantes confiáveis e consultáveis.

3 – Carregamento

O ETL move os dados transformados para o armazenamento de dados de destino. Esta etapa pode implicar o carregamento inicial de todos os dados de origem ou pode ser o carregamento de alterações incrementais nos dados de origem. Você pode carregar os dados em tempo real ou em lotes programados.

Para esta pesquisa a soma dos conceitos CRISP-DM que é uma modelagem, com o ETL que é um processo, foi possível conceber uma ferramenta que tem a capacidade de fornecer informações direcionadas para a necessidades do negócio. Ferramenta esta que foi concebida na plataforma EXCEL, da fabricante Microsoft. Esta ferramenta terá condições de apresentar informações visuais através de um DASHBOARD interativo para facilitar a sua utilização e entendimento.

Resultados e discussão

No período de pesquisa foi necessário buscar conhecimentos a respeito do Data Science, não seria possível conectar a modelagem CRISP-DM e o processo ETL sem o conhecimento de seus conceitos. A ferramenta só pode ser concebida com conhecimento básico da plataforma EXCEL, ferramentas do EXCEL, como, tabela dinâmica, fragmentação de dados e criação de gráficos dinâmicos só foram possíveis após pesquisas sobre a utilização da plataforma EXCEL.

Passou-se então para a busca por dados. Conseguiu-se os dados em um site especializado em treinamentos online. Esta empresa disponibiliza uma tabela contendo dados de vendas de produtos de uma empresa não citada na fonte de dados. Os dados são usados para estudos de caso e treinamentos. Tais dados permitiram a pesquisa a fazer um itinerário da modelagem CRISP-DM conectada ao processo ETL e gerar a ferramenta que era o objetivo da pesquisa. Os resultados se mostraram satisfatórios, pois geraram informações que podem balizar o negócio.



Ciência, Tecnologia e Inovação no Brasil

O DASHBOARD gerou informações como, total de vendas, total de vendas por produto, gráfico com total de vendas, formas de pagamento por região e UF e outro gráfico com a porcentagem de venda por região. Todas as informações podem ser filtradas por ano, período, região, UF e produto.

O DASHBOARD foi eficaz, pois utilizou os dados fornecidos, que até então não tinham a capacidade de fornecer informação alguma e gerou informações que podem balizar tal negócio.

Conclusões

Pode-se concluir nesta pesquisa que, uma simples modelagem aliada à um conhecimento prático do negócio pode gerar informações extremamente valiosas para os tomadores de decisão. Conclui-se que, o EXCEL pode ser grande aliado dos negócios, pela sua versatilidade de uso, uma ferramenta que já é extremamente usada para elaboração de relatórios e documentos essenciais para o funcionamento do negócio, também pode ser usado para criar uma ferramenta que trata dados e os conecta de maneira coesa para enfim fornecer informações valiosas. Esta Pesquisa demonstrou que, em um futuro muito próximo, as empresas que resistirem ao avanço do Data Science e seus conceitos, não resistiram ao mercado que hoje já é extremamente competitivo. Esta pesquisa apresentou uma ferramenta que busca solucionar esta resistência tendo em vista a sua simplicidade e também a costumeira familiaridade dos negócios com a plataforma EXCEL.

Agradecimento

Agradeço a Deus, a minha esposa e filhos além de meu orientador Professor Alexandre Pinho, que nos momentos de dificuldades e desânimo me foram o ombro amigo e consolador. Além da UNIFEI e CNPq que deram a oportunidade de me desafiar e de obter essa experiência maravilhosa.

Referências

SILVA, J. A. **Importância das informações contábeis para a tomada de decisões na administração pública municipal**. Revista Eletrônica Administração e Ciências Contábeis, v. 2, p. 1-12, 2009.

O'BRIEN, J. A. **Sistemas de Informação e as Decisões Gerenciais na Era da Internet**. Saraiva, 2002.

STAIR, Ralph. M; REYNOLDS, G. W. **Princípios de Sistemas de Informação**. Tradução da 11ª Edição Norte-Americana. Editora Cengage Learning. São Paulo, 2016.

KIMBALL, Ralph. **Data Warehouse Toolkit: Técnicas para construção de Data Warehouses dimensionais**. Editora MAKRON, 1998.

INMON, William H. **Como Construir o Data Warehouse**. 2ª ed. Rio de Janeiro: Editora Campus, 1997.

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

CHAPMAN, P. et al. **CRISP-DM 1.0: Step-by-step data mining guide**. SPSS inc, v. 9, p. 13, 200