



9º ENTEC – Encontro de Tecnologia: 23 a 28 de novembro de 2015

MINERAÇÃO DE DADOS APLICADO NO SETOR COMERCIAL E USO DE SOFTWARE PARA GESTÃO

Juliana Taise Cunha¹; Eduardo Chagas de Oliveira²

^{1,2} Instituto de Engenharia e Tecnologia – Universidade de Uberaba (UNIUBE)

juliana_eng.compy@hotmail.com; eduardo.chagas@uniube.br

Resumo

Muito tem se falado sobre o crescente aumento de informações nas bases de dados de empresas e organizações e da necessidade de investimento em inovação tecnológica, devido sobretudo, ao cenário atual, de um mercado bastante competitivo. Este trabalho apresenta um *software* de controle de vendas e estoque, desenvolvido para uma empresa do setor “varejo”, que não fazia uso de nenhum tipo de recurso tecnológico, e ainda, a aplicação da mineração de dados sobre a base real desta empresa, adquirida após um tempo de uso do *software*.

Palavras-chave: Mineração de Dados. Banco de dados. Conhecimento. Gestão.

1 Introdução

Gerenciar uma empresa é sem dúvida, uma tarefa que exige dos empresários muito conhecimento, responsabilidade e cautela para tomada de decisões, visto que uma simples falha gerada por um mal planejamento, pode trazer sérios riscos ou prejuízos para os mesmos (ANDRADE; AMBONI, 2011).

À medida que a empresa cresce, esse gerenciamento vai se tornando cada vez mais difícil, sendo necessário buscar por recursos tecnológicos que facilitem esse processo de gestão e aquisição de conhecimento.

O crescente aumento da procura desses recursos, somado a um aumento significativo de informações que vão sendo armazenadas nos bancos de dados das empresas e organizações, deram origem ao *Data Mining* (DM) ou Mineração de Dados. Trata-se de um processo que

se caracteriza por vasculhar esses bancos com grandes números de dados, utilizando diversos tipos de técnicas e algoritmos, definidos de acordo com cada aplicação, para abstrair informações valiosas, inicialmente imperceptíveis e transformá-las em conhecimento, que contribuirão fortemente, para a tomada de decisão (WITTEN; FRANK; HALL, 2011).

Dentre as várias técnicas de mineração de dados, podemos citar: regras de associação, que consiste em encontrar métodos ou regras do tipo $x \rightarrow y$, ou seja, (se x ocorre, então y ocorre), que costumam acontecer com determinada frequência (TAN; STEINBACH; KUMAR, 2009) e árvore de decisão, modelo com características de uma árvore contendo raiz e nós (itens que aparecem abaixo da raiz), que pode ser usado para prever valores para o futuro, com base em atributos conhecidos, de forma a subdividir um problema em outros menores, chamados de nós filhos, até que se encontre uma solução para cada problema, chegando então aos nós folhas (GONZALES; ZAMPIROLI, 2014).

Atualmente, o estudo da Mineração de Dados tem obtido espaço significativo no campo de pesquisa, devido aos bons resultados já obtidos. Dessa forma, torna-se cada vez mais interessante aprofundar-se nessa área de estudo, buscando colaborar com futuros pesquisadores de forma direta ou indireta, por meio de análises que levem a percepção de possíveis aperfeiçoamentos e/ou descoberta de novas áreas onde possa ser aplicado.

www.uniube.br/entec - UNIUBE Campus Aeroporto – Uberaba/MG

9º ENTEC – Encontro de Tecnologia: 23 a 28 de novembro de 2015

O presente trabalho tem por objetivo realizar um estudo sobre Mineração de Dados e utilizar as técnicas de associação e árvores de decisão para minerar a base de dados de um supermercado, onde os dados serão armazenados por meio da utilização de um *software* de gestão, também desenvolvido para esta empresa, buscando assim, agilizar o atendimento e potencializar os processos de decisão da mesma.

2 Materiais e Métodos

Este trabalho trata-se de um estudo de caso, realizado em uma empresa do setor varejista (Supermercado Pingo de Ouro), situado na cidade de Iraí de Minas, em que a análise e coleta de informações foi realizada de forma exploratória, através de visitas ao local e troca de ideias com o proprietário e o desenvolvimento foi prático.

2.1 Estudo de caso

Em análise ao status da empresa investigada, foi verificado no momento de análise, a necessidade de investimento em inovação tecnológica, visto que ela ainda trabalhava com o padrão antigo, em que os clientes passam as compras no caixa e a única ferramenta disponível é a calculadora e o controle de vendas é feito por meio de anotações. Além disso, verificou-se a necessidade de encontrar padrões que permitissem ao proprietário, ter um maior conhecimento de sua empresa e seus clientes.

Inicialmente foi desenvolvido um *software* que pudesse atender à primeira necessidade, e após 3 meses de uso do *software*, foi aplicado à base de dados da empresa, o processo de mineração de dados.

2.2 Desenvolvimento do Software

O *software* foi desenvolvido na linguagem de programação orientada a objetos VB.NET, por meio da ferramenta

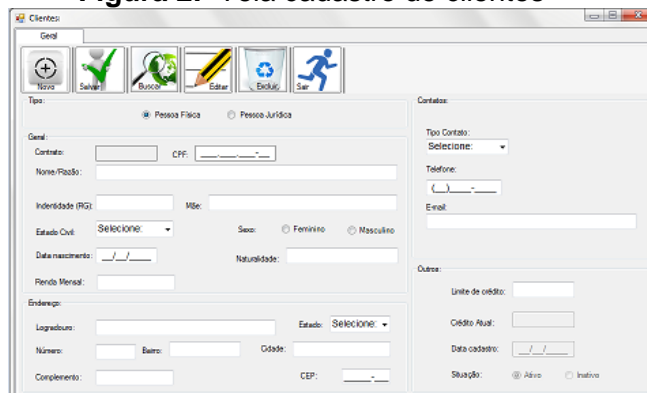
de desenvolvimento Visual Studio 2010, integrado ao sistema gerenciador de banco de dados SQL SERVER 2008, onde são armazenados os dados.

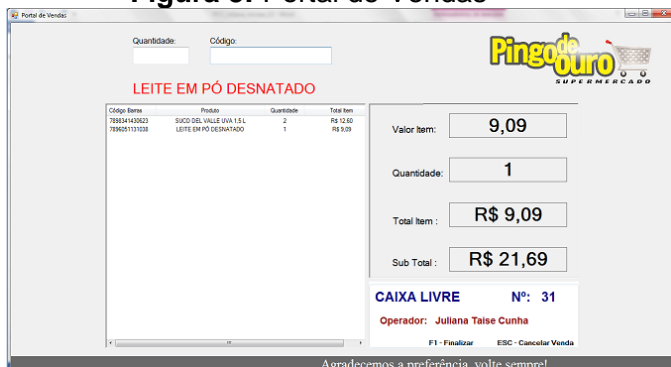
O sistema desenvolvido permite realizar todo o controle de vendas e estoque da empresa e conta com diversas funcionalidades, tais como: tela de *login* para segurança do sistema, mostrado na Figura 1, em que é obrigatório a troca de senha após primeiro acesso do usuário; tela de produtos, usuários, categorias, clientes (Figura 2), com opções para cadastro, consulta, alteração e exclusão, e liberados de acordo com o perfil de cada usuário, que pode ser (administrador ou operador); tela de lançamentos de produtos; portal de vendas (Figura 3); tela para troca de senha; relatórios e *logoff* do sistema.

Figura 1: Tela de Login



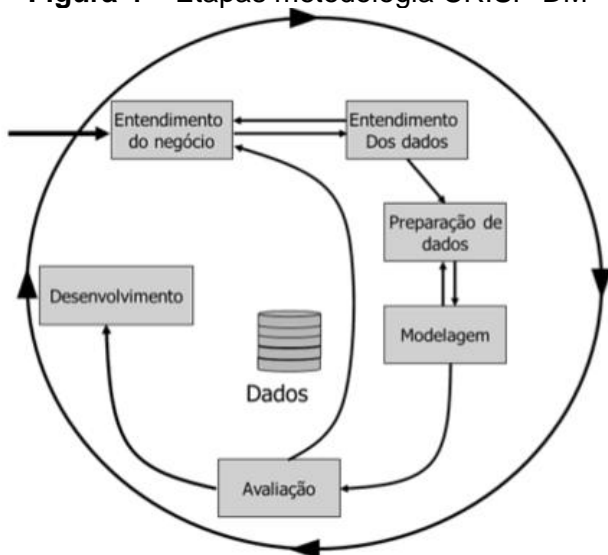
Figura 2: Tela cadastro de clientes



9º ENTEC – Encontro de Tecnologia: 23 a 28 de novembro de 2015
Figura 3: Portal de Vendas

2.3 Mineração de Dados

Para aplicar o processo de mineração sobre a base de dados foi utilizado a ferramenta RapidMiner que implementa um ou mais algoritmos para cada técnica de mineração, como exemplo o algoritmo Apriori, usado para trabalhar com regras de associação e ID-3 para árvores de decisão.

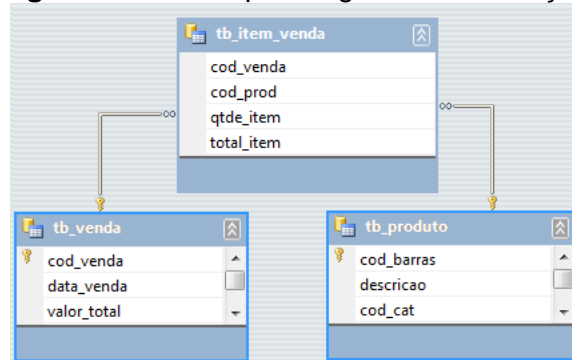
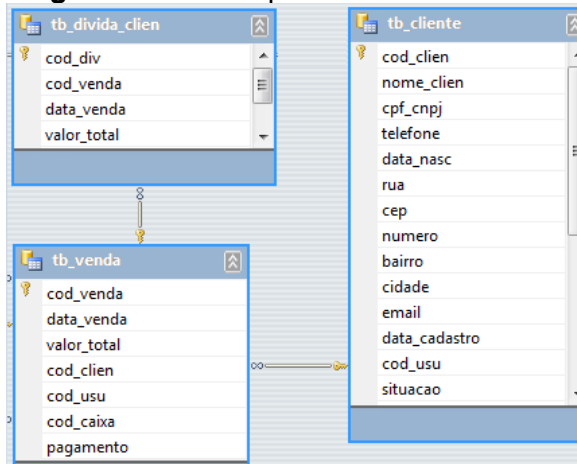
Utilizou-se ainda a metodologia CRISP-DM, modelo padrão de mineração, criado em 1999 (NORTH, 2012). A Figura 4, representa as etapas desta metodologia.

Figura 4 – Etapas metodologia CRISP-DM


A seguir é detalhado como o processo foi desenvolvido, de acordo com as 6 etapas.

1- Entendimento do Negócio: A ideia foi buscar padrões que permitissem ao proprietário, conhecer as preferências de seus clientes, possibilitando a ele, fazer um melhor planejamento das compras. Além disso, encontrar padrões que pudessem beneficiar no aumento de vendas e prever se um grupo de clientes que possui dívidas em aberto no estabelecimento tende ou não a quitá-las.

2- Entendimento dos dados: Buscou-se nesta etapa, entender o significado dos dados representados em cada tabela do banco de dados da empresa. A Figura 5 mostra a estrutura das tabelas utilizadas para regras de associação. Já a Figura 6, representa a estrutura das tabelas usadas para trabalhar com árvores de decisão.

Figura 5: Tabelas para regras de associação

Figura 6: Tabelas para árvore de decisão


9º ENTEC – Encontro de Tecnologia: 23 a 28 de novembro de 2015

- 3- Preparação dos dados: Para iniciar o processo de mineração de dados é necessário que não existam dados inconsistentes (campos com valores nulos) ou em formatos inadequados, ou seja, atributos com tipo de dados diferentes do formato aceito em cada técnica de mineração, assim as amostras com valores nulos foram eliminadas, visto que ocorreram em pequena quantidade se comparado com a quantidade total de amostras e os atributos foram selecionados e colocados nos formatos adequados.
- 4- Modelagem: Foi feita a montagem do modelo, que representa todo o processo de mineração, através do uso de operadores disponíveis na ferramenta de mineração, posteriormente cada processo foi executado. A Figura 7 mostra o modelo usado para a técnica de regras de associação e a Figura 8 para a técnica de árvore de decisão.

- 5- Avaliação: Nesta etapa foi realizada uma análise, para entendimento dos resultados, sempre comparando se poderiam beneficiar de alguma forma, ao que foi apresentado na primeira etapa. Para os resultados obtidos na regra de associação foi realizado novos testes, mudando os parâmetros de suporte e confiança para obtenção de melhores resultados.
- 6- Desenvolvimento/ Apresentação: Foi realizada a montagem de gráficos e imagens que pudessem permitir uma melhor visualização e entendimento dos resultados, que serão apresentados na seção 3.

3 Resultados

Por meio do uso de regras de associação, com parâmetros de suporte mínimo igual a 1% e confiança mínima igual a 75%, foi possível identificar dois padrões de consumo dos clientes, conforme apresentado abaixo:

1º- Um cliente que leva o produto “tinta para cabelo” tende a levar o produto “esmalte” na mesma compra, e esta regra possui confiança igual a 82%.

2º- Da mesma forma, um cliente que leva o produto “macarrão lasanha”, tende a levar na mesma compra, o produto “Coca-Cola”, porém esta regra possui confiança igual a 75%.

As Figuras 9 e 10 representam os dois padrões de consumo encontrados.

Figura 7: Modelo para regras de associação

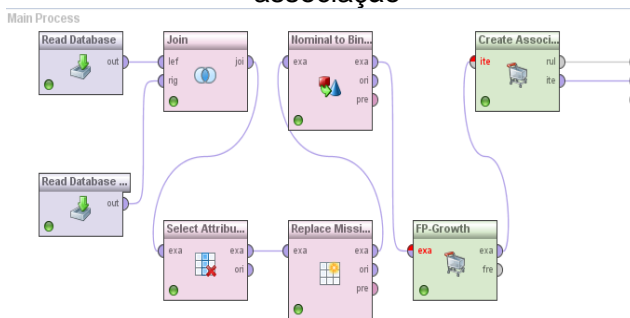


Figura 8: Modelo para árvore de decisão

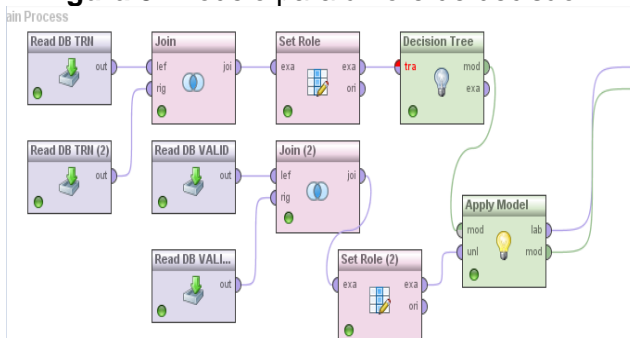


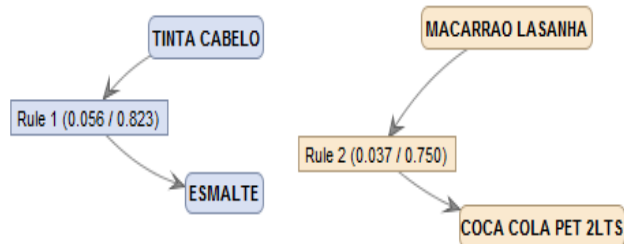
Figura 9: Padrões de consumo

Association Rules

```
[MACARRAO LASANHA] --> [COCA COLA PET 2LTS] (confidence: 0.750)
[TINTA CABELO] --> [ESMALTE] (confidence: 0.823)
```

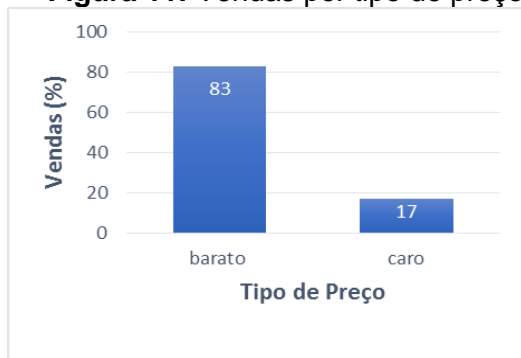

9º ENTEC – Encontro de Tecnologia: 23 a 28 de novembro de 2015

Figura 10: Fluxo padrões de consumo



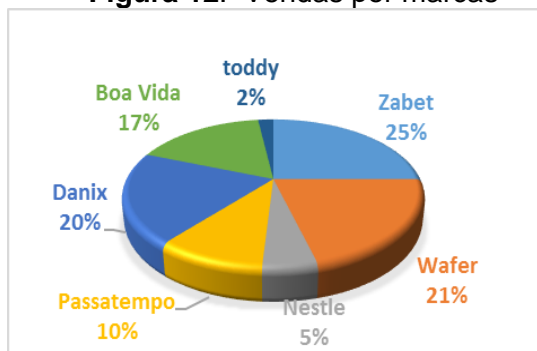
Outro padrão encontrado por meio dessa técnica foi a preferência dos clientes por algumas marcas de produtos da categoria biscoitos recheados, na qual as marcas disponíveis eram passatempo, todody, wafer, zabet, danix, boa vida e nestle. Foi então identificado que em 83% das vendas realizadas que possuíam este tipo de produto, os clientes tiveram preferência pelas marcas mais baratas, como mostrado na Figura 11.

Figura 11: Vendas por tipo de preço



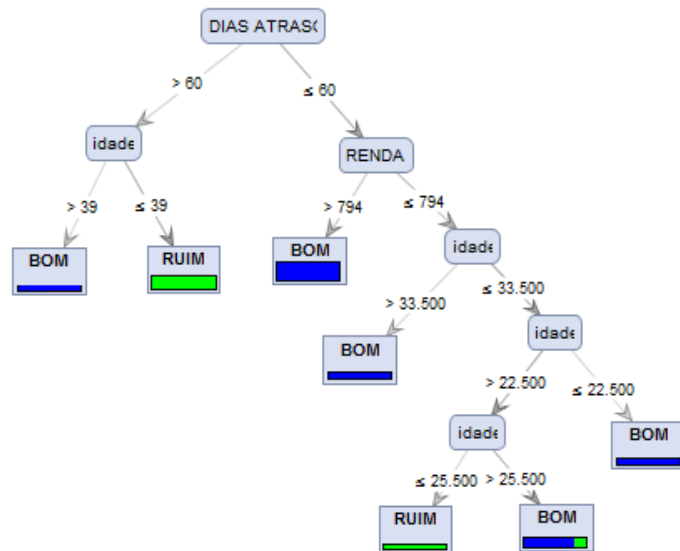
A Figura 12 mostra ainda, a relação da quantidade de vendas realizadas em porcentagem, para cada marca dos produtos desta categoria.

Figura 12: Vendas por marcas



Por meio da técnica de árvore de decisão foi possível prever que para os clientes analisados, aqueles que possuem dívidas com mais de 60 dias de atraso e idade menor ou igual a 39 e aqueles com dívidas que tenham 60 dias de atraso ou menos, renda menor que 794 reais e idade menor que 25, tende a não quitar suas dívidas. Além disso, para cada cliente analisado foi obtido um grau de confiança para o resultado encontrado. A Figura 13 representa a estrutura da árvore de decisão obtida, onde “bom” representa os clientes que tendem a pagar e “ruim”, os clientes que tendem a não pagar.

Figura 13: Árvores de decisão



4 Discussão

Os padrões identificados geraram novos conhecimentos que antes eram desconhecidos e que podem ser altamente importantes para a tomada de decisões dentro desta empresa.

Visto que o proprietário tinha grande perda com os produtos da categoria biscoitos recheados, a aquisição de conhecimento, com relação à preferência de seus clientes por este tipo de produto, dará uma nova noção de planejamento para compra dos produtos desta categoria, onde ele pode até mesmo não

9º ENTEC – Encontro de Tecnologia: 23 a 28 de novembro de 2015

mais comprar aquelas marcas que tiveram baixo percentual de aceitação.

Tendo também o conhecimento de quais clientes tem maior possibilidade de não pagar suas dívidas, os resultados poderão influenciar em várias tomadas de decisões, como exemplo, cortar o crédito desses clientes para evitar perdas maiores.

Além disso, o proprietário poderá utilizar a relação encontrada dos produtos que costumam ser vendidos na mesma compra para fazer propagandas, promoções, sugestões ao cliente, definição do local de disponibilização dos produtos, como forma de chamar mais a atenção dos clientes.

O uso do *software* poderá também trazer novos clientes para a empresa, visto que o atendimento se tornará mais rápido, além de facilitar o controle e gestão da mesma.

5 Conclusão

O processo de mineração de dados pode ser utilizado para descobrir diversos padrões importantes. No entanto, para utilizá-lo e obter resultados que realmente possam gerar novos conhecimentos é essencial seguir à risca todo um fluxo para entendimento, preparação dos dados e análise dos resultados, na qual todas as etapas possuem alto grau de importância dentro do processo. Além disso, é preciso fazer o uso de técnicas adequadas, que podem variar de acordo com o que é esperado como resultado.

Tendo estes resultados em mãos, cabe à pessoa que estava à procura desses conhecimentos, saber utilizá-los, tomando decisões adequadas, que possam trazer de alguma forma, algum retorno satisfatório e benéfico à empresa.

Referências

ANDRADE, Rui Otávio Bernardes de; AMBONI, Nério. **Teoria geral da administração**. 2ª Ed. Rio de Janeiro: Elsevier-Campus, 2011. 288p

GONZALES, José Artur Quilici; ZAMPIROLI, Francisco de Assis. **Sistemas Inteligentes e Mineração de Dados**. São Paulo: Triunfal Gráfica e Editora, 2014. 148 p.

NORTH, Matthew. **Data Mining for the Masses**, 2012. Disponível em: <<https://rapidminer.com/wp-content/uploads/2013/10/DataMiningForTheMasses.pdf>>. Acesso em: 15 Setembro 2015.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Data Mining (Mineração de Dados)**. Rio de Janeiro: Editora Ciência Moderna Ltda., 2009. 978 p.

WITTEN, Ian; FRANK Eibe; HALL Marck. **Data Mining: Practical Machine Learning Tolls and Techniques**. Second Edition. Amsterdam: Morgan Kaufmann Publisher, 2011. 629 p.