

## **Avaliação e Análise de Softwares para Mineração de Dados aplicados ao processo de Pós-Compra**

**Haroldo G. B. Filho<sup>1</sup>, Raimundo C. Soares<sup>1</sup>, Louremar V. A. Júnior<sup>1</sup>, Elias S. Silva<sup>1</sup>**

<sup>1</sup>Coordenação de Sistemas para Internet – Faculdade de Educação de Bacabal (FEBAC)  
Rua 12 de Outubro, 377, Centro, Bacabal - MA, 65700-000

{haroldogomes86}@gmail.com

***Abstract.** In an increasingly competitive market and business models enhance become increasingly faster, where the post purchasing process is characterized as a critical decision-making step for a customer loyalty and thus ripening of the mark. To enable companies to achieve this phase successfully proposed the use of software focused on data mining, for compression and refinement of information in relation to a particular customer profile, but it is valuable to assess and analyze which software is best suited for each scenario.*

***Resumo.** Em um mercado cada vez mais competitivo, os modelos de negócio aprimoram-se cada vez mais rápido, onde o processo de pós-compra caracteriza-se como uma fase de tomada de decisão crucial para a fidelização de um cliente e conseqüentemente amadurecimento da marca. Para que as empresas possam alcançar tal fase com êxito é proposto a utilização de softwares focados em Mineração de Dados, para a compressão e refinamento de informações em relação a um determinado perfil de cliente, porém é válido avaliar e analisar qual software é melhor indicado para cada cenário.*

### **1. Introdução**

Os processos de desenvolvimento de produtos, aplicativos de celular, *softwares*, entre outros, vem evoluindo progressivamente, porém torna-se irrelevante vender algo sem “fidelizar” o consumidor. Devido a esse fato de preocupa-se com a satisfação do cliente não apenas na etapa do consumo, mas após ele, investisse em modelos matemáticos, softwares e técnicas baseadas em Mineração de Dados para o entendimento de novas e complexas estruturas de dados oriundos do volume de informação e transformações em padrões de compra.

### **2. Processo de Pós-Compra**

O Processo de Pós-compra por se encontrar no fim de um ciclo de tomadas de decisão envolvendo o sentimento e comportamento do cliente, baseia-se em um fluxo regular de relacionamento, para [Polloni et al 2004] é necessário considerar aspectos como: propósito, relacionamento e satisfação, para que tal ciclo tenha por consequencia o retorno do consumidor.

O propósito de todo negócio, bens ou serviços, é criar e manter clientes satisfeitos [Drucker 1973]. Um dos grandes atributos existentes no sucesso de um processo de compra está relacionada ao relacionamento duradouro entre empresa e consumidor (Rapp; Collins, 1990 – *apud* Engel, Blackwell, Miniard, 2000). Para que este relacionamento mantenha-se ao longo do tempo é preciso entender a necessidade da existência de qualidade e satisfação permanente dos clientes. O foco na qualidade, do ponto de vista do cliente, e sua satisfação são fundamentais para o sucesso do processo de compra. (Lovelock, Wight, 2001)

A Satisfação de compra é o resultado da diferença entre desempenho percebido pelo comprador e suas expectativas anteriores de compra de determinado produto ou serviço [Kotler 2000]. Segundo Mowen e Minor (2003), a satisfação é “a atitude geral sobre um produto ou serviço após a sua aquisição e uso. é o julgamento de avaliação posterior à compra, resultante de uma compra específica” .

Deve-se considerar que uma vez feita a venda, o processo de avaliação de satisfação e qualidade estende-se em especial à fase de Pós-Compra, que será o momento que o consumidor poderá efetivamente evidenciar o desempenho da compra obtida. (Engel, Blackwell, Miniard, 2000)

### **3. Mineração de dados**

Com o aumento do armazenamento de dados crescendo exponencialmente, algumas questões tornaram-se relevantes: Como tratar tais dados ? Como as bases de dados suportarão tamanha demanda ? Existem técnicas e *softwares* específicos para tais situações ? Com o objetivo de responder tais perguntas, foi proposta, no final da década de 80, a Mineração de Dados, do inglês *Data Mining*.

A Mineração de Dados é uma das tecnologias que mais vem crescendo ao longo da última década, em detrimento do capital investido por companhias com o objetivo de coletar dados com precisão e refinamento. Witten et al. [88], Olson et al. [58] e Bramer [6] apresentam algumas das aplicações em um fluxo de compra nas quais a Mineração de Dados é aplicada de forma eficaz:

Retenção de clientes: identificação de perfis para determinados produtos, venda cruzada;

Bancos: identificar padrões para auxiliar no gerenciamento de relacionamento com o cliente;

Cartão de Crédito: identificar segmentos de mercado, identificar padrões de rotatividade;

Cobrança: detecção de fraudes;

Telemarketing: acesso facilitado aos dados do cliente.

Por possuir um caráter multidisciplinar, a Mineração de Dados adapta seu conceito ao cenário de atuação de cada autor, destacando-se aplicações em: Estatística, Aprendizado de Máquina e Banco de Dados. Em [Zhou 1996], é feita uma análise comparativa sobre as três perspectivas citadas.

Em [Hand et al 2001], a definição é dada de uma perspectiva estatística: "Mineração de Dados é a análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados".

Em [Cabena et al 1998], a definição é dada de uma perspectiva de banco de dados: "Mineração de Dados é um campo interdisciplinar que junta técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados".

Em [Fayyad et al 1996], a definição é dada da perspectiva do aprendizado de máquina: "Mineração de Dados é um passo no processo de Descoberta de Conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que, sob certas limitações computacionais, produzem um conjunto de padrões de certos dados."

Habitualmente a Mineração de Dados é identificada como uma tecnologia capaz de realizar tarefas complexas, tais como:

### **3.1. Descrição**

É a tarefa utilizada para descrever padrões e tendências demonstrados por dados. Em geral, uma descrição promove uma gama de possibilidades para possíveis interpretações de resultados obtidos, bastante utilizada em fases de análise de dados.

### **3.2. Classificação**

Possui o objetivo de identificar a qual classe um determinado registro pertence. o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de 'aprender' como classificar um novo registro (aprendizado supervisionado) [Polloni et al 2004].

### **3.3. Estimação**

Tal tarefa é similar à classificação, no entanto é utilizada quando um registro é identificado por um valor numérico e não um categórico. Possui um foco quantitativo para avaliação, por exemplo, de acordo com um número de compradores de um determinado produto, dizer quanto foi investido, baseado em tal informação pode-se lestimar qual produto poderá dar um certo prejuízo em alguns meses ou dias.

### **3.4. Predição**

A tarefa de predição é similar às tarefas de classificação e estimação, porém ela visa descobrir o valor futuro de um determinado atributo.

### **3.5. Agrupamento (*Clustering*)**

A tarefa de agrupamento visa identificar e aproximar os registros similares. Um agrupamento (ou *cluster*) é uma coleção de registros similares entre si, porém diferentes

dos outros registros nos demais agrupamentos. Esta tarefa difere da classificação pois não necessita que os registros sejam previamente categorizados (aprendizado não-supervisionado). Além disso, ela não tem a pretensão de classificar, estimar ou prever o valor de uma variável, ela apenas identifica os grupos de dados similares [Polloni et al 2004].

### 3.6. Associação

Através da forma: SE atributo  $X$  ENTÃO atributo  $Y$ , identifica a relação entre grupos através de análise de dados, aplicado a áreas como: Segurança, *Market Basket*, Taxonomia de plantas, entre outras.

## 4. Softwares Para Mineração De Dados

Para uma descoberta de conhecimento, bem como o seu respectivo refinamento, existem um quantidade expressiva no mercado, com áreas de atuação diferentes, livres ou “pagos”, complexos ou simples, aplicados em mineração de dados para formalização de perfil, mineração de texto, numérica, processamento de imagens, entre outras. Para a avaliação e análise em processo de pós-compra (objeto de pesquisa do presente trabalho) selecionou-se alguns softwares com alta participação de mercado, são eles: Weka, Oracle Data Mining (ODM), IBM SPSS Modeler e GMDH *Shell*.

### 4.1. WEKA

*Waikato Environment for Knowledge Analysis* – WEKA [Waikato 2004, Witten & Frank 2000] é uma ferramenta de KDD (*Knowledge Discovery in Databases*, ou seja, Descoberta de Conhecimento em Bancos de Dados) que contempla uma série de algoritmos de preparação de dados, de aprendizagem de máquina (mineração) e de validação de resultados [Silva 2004]. WEKA foi desenvolvido na Universidade de Waikato na Nova Zelândia, sendo escrito em Java e possuindo código aberto disponível na Web (ver figura 1), fornecendo funcionalidades para pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização, atualmente faz integra a ferramenta *BI OpenSource Pentaho*.

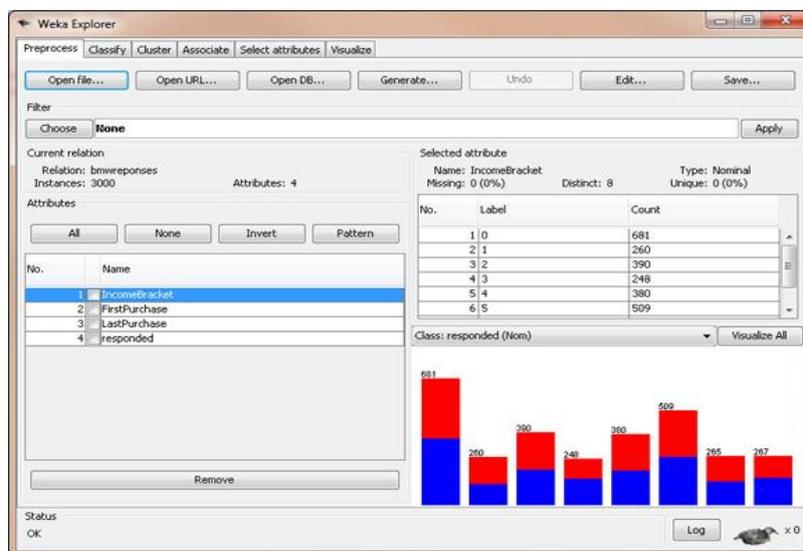


Figura 1. Interface do Weka.

## 4.2. Oracle Data Mining (ODM)

O Oracle Data Mining (ver figura 2) integra o pacote de soluções para base de dados conhecido como Oracle Database 11g Enterprise Edition, onde disponibiliza técnicas para encontrar padrões, prever comportamentos, identificar atributos chave e associações escondidas ou valores anormais. Possui recursos específicos para gerenciamento de desempenho de algoritmos, para que o mesmo possa utilizar o método *dividir para conquistar* frente a situações mais complexas. Suporta técnicas de *supervised* e *unsupervised learning* e inclui algoritmos de Classificação e Regressão, Modelos de *Clustering* e Associações. O ODM é utilizado através de interfaces em SQL (*Structured Query Language*), PL/SQL, Java e Add-Ins para *Excel*

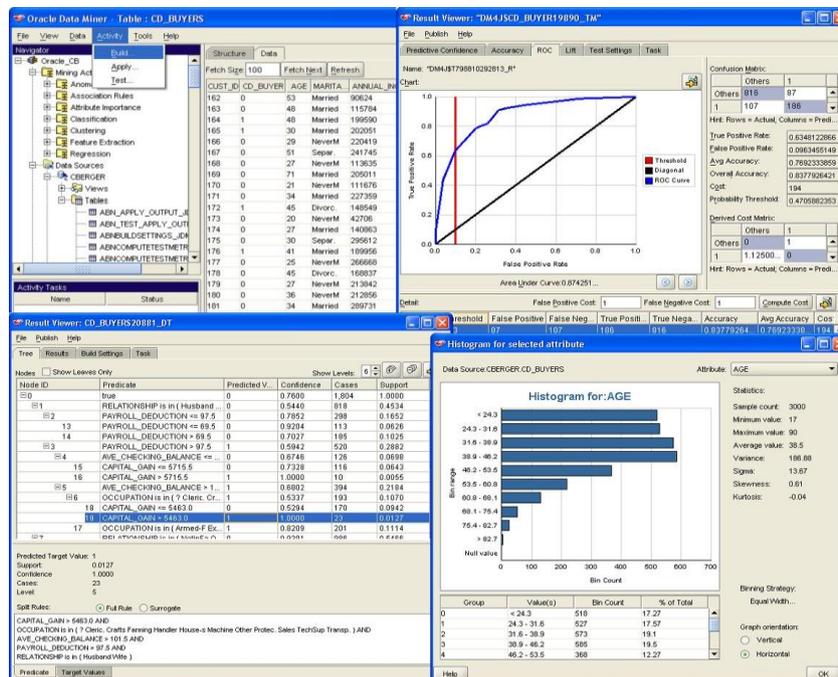


Figura 2. Multi-interface do Oracle Data Mining

## 4.3. IBM SPSS Modeler

O IBM SPSS *Modeler* (ver figura 3) é um software pertencente a IBM (Internacional Business Machines) caracteriza-se por ser uma plataforma de análise de predição que foi desenvolvida para modelar e representar a inteligência em futuras tomadas de decisão por grupos, indivíduos ou sistemas. Fornece vários algoritmos avançados e técnicas que incluem análise de texto, análise de entidades, gestão de decisões e otimização.

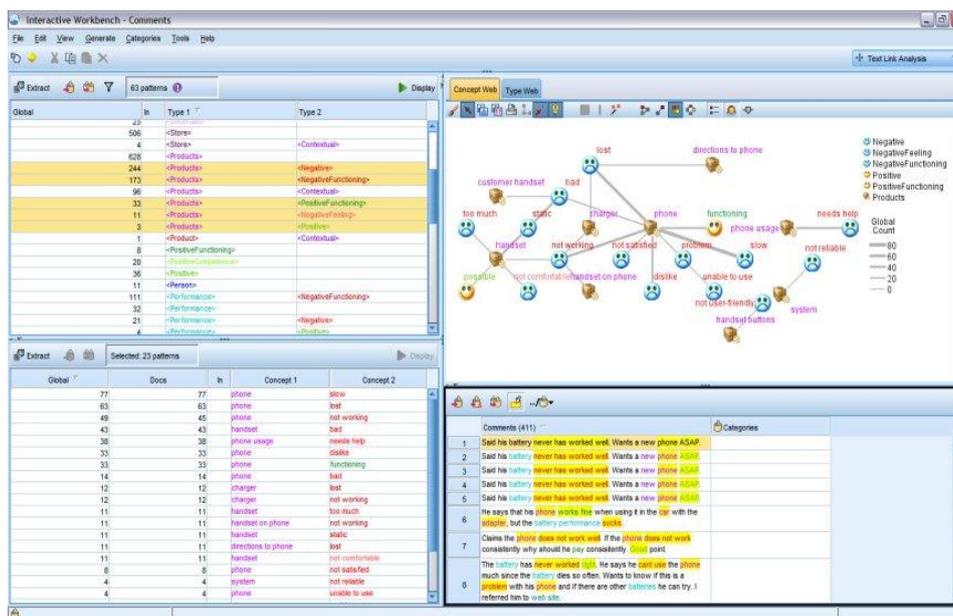


Figura 3. Interface do IBM SPSS Modeler

#### 4.4. GMDH Shell

GMDH Shell (ver figura 4) é uma ferramenta *easy-to-use* para mineração de dados e previsão de conjuntos de dados multi-paramétricos. Ele executa uma otimização estrutural e paramétrica totalmente automático de um modelo, é útil para mineração de dados, análise preditiva, análise de séries temporais, a previsão e descoberta de conhecimento. O *software* combina a tecnologia de aprendizado de máquina bem comprovada e capacidades alargadas para o uso eficaz de multi-core, com múltiplos processadores e computadores em *cluster*.

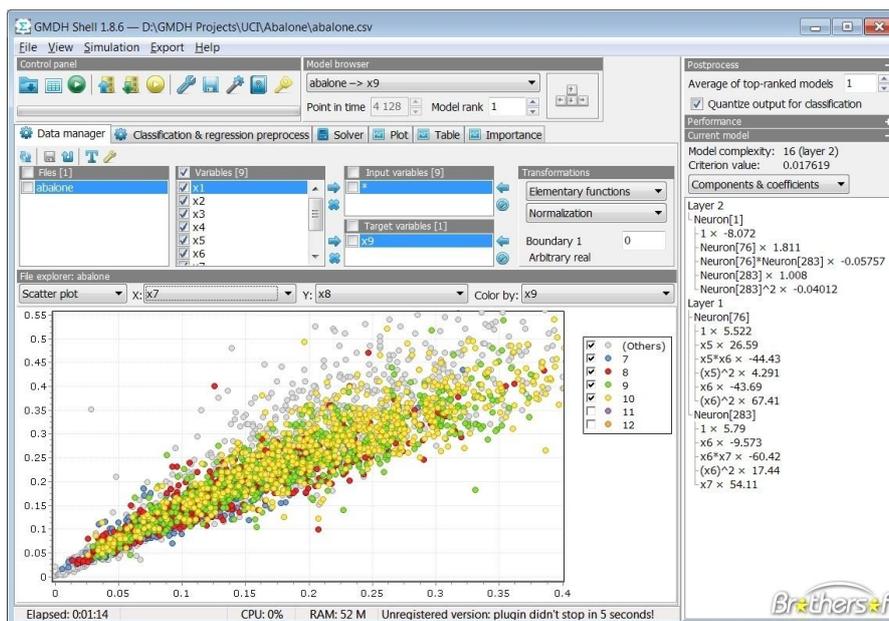


Figura 4. Interface do GMDH Shell

GMDH Shell torna o processamento de dados muito mais fácil, ele é capaz de detectar automaticamente os dados utilizáveis dentro de um arquivo, transformar os dados de acordo com um tipo de problema, descartar entradas irrelevantes e construir um conjunto de modelos preditivos na base de princípios de detecção de complexidade e auto-organização de objetivos.

## 5. Metodologia para Avaliação e Análise

Para formalizar uma avaliação e análise coesa e com resultado minimamente errôneo necessitam-se de parâmetros, definidos para fins de tais testes: (I) Escalabilidade, a capacidade de evolução do software em função do tempo e sua robustez; (II) Usabilidade, disposição das funcionalidades de um *software*; (III) Curva de aprendizagem, nível de adaptação do usuário ao software; (IV) Extensões, se há funcionalidades adicionais em constante desenvolvimento; (V) Licença, participação de alguma comunidade ou anuidades; (VI) Documentação, embasamento teórico em função do *software* em questão, através de *helps*, artigos e patentes.

Para a validação dos parâmetros, e posteriormente os resultados, representa-se o seguinte estudo de caso: dez lojas eletrônicas possuem uma carteira de mil clientes, cada, dos quais duzentos não voltaram a comprar no site, trezentos não visitam o site e quinhentos mantiveram algum tipo de relacionamento. Vinte usuários simularam tal evento em quatro softwares (objetos de estudo deste artigo) e aplicaram “conceitos” de 1 (pior conceito) a 5 (melhor conceito), onde pode-se observar (ver tabela 1) tais resultados das médias:

	Escalabilidade	Usabilidade	Curva de aprendizagem	Extensões	Licença	Documentação
Weka	3,5	4,5	5	4	5	4,5
ODM	3	4	4	4	3	5
IBM SPSS	5	4	4	4,5	3	5
GMDH Shell	4	4,5	3,5	3	3,5	3,5

**Tabela 1. Médias de avaliação baseada em parâmetros**

Baseado em tais testes utilizando parâmetros, seis ao todo. Percebe-se que o *software* Weka possuiu um resultado mais satisfatório que os demais, porém é importante considerar que tal desempenho limitou-se ao estudo de caso com foco em um processo de pós-compra, não levando em consideração qualquer técnica ou algoritmo de mineração de dados, utilizando estritamente recursos e eventos gráficos (GUI, *Graphical User Interface* ou Interface Gráfica com usuário) ou “*click-to-result*”, ou seja, clique e tenha a resposta.

## 6. Conclusão

Percebe-se que tais softwares possuem um desempenho satisfatório, porém possuem certas limitações em requisitos cruciais, principalmente na Licença, que impacta diretamente na participação de uma comunidade ativa ou não, que conseqüentemente enfraquece ou limita a escalabilidade e documentação. É importante ponderar que os softwares aqui avaliados não se limitam exclusivamente ao cenário de um processo de pós-compra, mesmo que baseado nos resultados seja uma importante consideração.

## Referências

- Bramer, M. **Undergraduate Topics in Computer Science - Principles of Data Mining**. Springer, 2007.
- Cabena, P; Hadjinian, P; Stadler, R; Jaapverhees; Zanasi, A. **Discovering Data Mining: From Concept to Implementation**. Prentice Hall, 1998.
- Drucker, P. F. Management: tasks, responsibilities, practices. New York: Harper & Row, 1973.
- Engel, J. F., Blackwell, R. D., Miniard, P. W. Comportamento do consumidor. Rio de Janeiro: LTC, 2000. 8ª edição.
- Fayyad, U; Piatetsky-Shapiro, G; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. American Association for Artificial Intelligence, 1996.
- Hand, D; Mannila, H; Smyth, P. **Principles of Data Mining**. MIT Press, 2001.
- Lovelock, C.; Wright, L. Serviços: marketing e gestão. São Paulo: ed. Saraiva, 2001.
- Mowen, J. C.; Minor, M. S. Comportamento do consumidor. São Paulo: Prentice Hall, 2003.
- Polloni A., Barbieri L., Macedo A. L. A., Souza C. C. M., Maciel J. D., Riccardi R. Q., Garcia M. N., FECAP, 2002. O Comportamento do Consumidor no Pós-Compra – Identificando as Reclamações – Um Estudo Exploratório.
- Olson, D. L; Delen, D. **Advanced Data Mining Techniques**. Springer, 2008
- Silva M. P. S. Mineração de Dados - Conceitos, Aplicações e Experimentos com Weka, Trabalho de Conclusão de Curso, Universidade Estadual do Rio Grande do Norte, UERN.
- University of Waikato. Weka 3 – Machine Learning Software in Java. Disponível no site da University of Waikato (2004). URL: <http://www.cs.waikato.ac.nz/ml/weka>
- Zhou, Z.-H. **Three perspectives of data mining**. Artificial Intelligence Journal, p. 139–146, 2003.
- Kotler, P. Administração de Marketing. São Paulo: ed. Prentice Hall, 2000. 10ª edição.
- Witten, I.; Frank, E. Data Mining – Practical Machine Learning Tools. Morgan Kaufmann, 2000.