



MINERAÇÃO DE ÁRVORE DE DECISÃO PARA IDENTIFICAR PERFIS DE CIRCUNSTÂNCIAS DE ÓBITOS EM DADOS DO SIM

EICH, Stéfano Costa¹; SCHUCH, Regis Rodolfo²; CHICON, Patricia M.M³

Palavras-Chave: Informações. Departamento. Disseminar. Decisões.

INTRODUÇÃO

Este resumo é parte integrante de um trabalho de conclusão de curso que objetiva utilizar tecnologias de Banco de Dados como a integração da Mineração de Dados (MD) apoiando o processo de tomada de decisão aos gestores de Saúde Pública do Sistema Único de Saúde (SUS). Neste contexto, utilizou-se técnicas computacionais baseadas em Árvores de Decisão, amplamente utilizada em algoritmos de classificação, com o objetivo de extrair conhecimento das bases de dados do Departamento de Informática do SUS (DATASUS). Ao longo do trabalho será abordada a MD através da técnica de Árvores de Decisão e do Algoritmo J48, bem como os resultados obtidos através de sua aplicação sobre base de dados do Sistema de Informação Ambulatorial (SIM) do DATASUS.

MINERAÇÃO DE DADOS

Ao longo do tempo a maioria das operações e atividades das instituições privadas e públicas passaram a ser registradas computacionalmente e acumuladas em grandes bases de dados. O Descoberta de Conhecimento em Bases de Dados (DCBD) pode ser visto como o processo da descoberta de novas correlações, padrões e tendências significativas por meio da análise minuciosa de grandes conjuntos de dados estocados. Este processo se vale de tecnologias de reconhecimento utilizando padrões e técnicas estatísticas e matemáticas. (NORTON, 1999). A técnica da MD é uma das alternativas mais eficazes para extrair conhecimento a partir de grandes volumes de dados, descobrindo relações ocultas, padrões e estabelecendo regras para predizer e correlacionar dados, que podem auxiliar as instituições nas tomadas de decisões mais rápidas ou, até mesmo, a atingir um maior grau de assertividade. (AVELAR, 2017). Conforme (CREPALDI, 2012):

¹ Discente do Curso de Ciência da Computação. Unicruz. E-mail: ste.colorado@gmail.com

² Professor, UNICRUZ. E-mail: regis.schuch@gmail.com

³ Professora Ciência da Computação na Universidade de Cruz Alta - UNICRUZ. E-mail: pmozzaquatro@unicruz.edu.br



Árvores de Decisão são métodos de classificação de dados no contexto da chamada MD. Podem ser usadas em conjunto com a tecnologia de indução de regras, mas são as únicas a apresentar os resultados hierarquicamente (com priorização).

O algoritmo J48 é uma implementação em Java do algoritmo C4.5 (QUINLAN, 1993) que gera árvores de classificação a partir de um conjunto de dados de treinamento, sendo assim, a cada nó, o algoritmo escolhe um atributo que mais eficientemente subdivide o conjunto das amostras em subconjuntos homogêneos e caracterizados por sua classe. O critério é o ganho de informação obtida na escolha do atributo para subdivisão (QUINLAN, 1993). A principal característica das Árvores de Decisão é a tomada de decisões levando em consideração os atributos mais relevantes, além de ser mais concebível para um maior número de pessoas. Ao escolher e apresentar os atributos em ordem de importância, a Árvore de Decisão possibilita aos usuários conhecer quais fatores mais influenciam os seus trabalhos.

METODOLOGIA E DESENVOLVIMENTO

Para o desenvolvimento do trabalho foi utilizado como recursos: o Sistema Gerenciador de Banco de Dados (SGBD) IBM DB2, produzido pela International Business Machines (IBM); e, como ferramenta para aplicação da MD, a Waikato Environment for Knowledge Analysis (WEKA), que contribuiu para organizar e aplicar técnicas sobre os algoritmos escolhidos para o cenário do trabalho. Ambos os recursos utilizados apresentaram uma versão gratuita.

O desenvolvimento da pesquisa consiste na Mineração dos dados do SIM do DATASUS, que foi realizado em várias etapas. Essas etapas são compostas do processo de DCBD, que de maneira coordenada permitiu extrair conhecimentos da base de dados. O processo é descrito a seguir:

- Na fase de seleção é escolhido o conjunto de dados contendo os atributos e características. Para uma boa análise desses registros, a fase de seleção separa os atributos e características de acordo com as informações de entrada.
- Na etapa de Pré-processamento foram eliminados dados redundantes e dispensáveis para não acumular informações desnecessárias. Também foram usados métodos para reduzir o número de variáveis no processo, melhorando o desempenho do algoritmo.
- Na fase que antecede a fase de MD, acontece a etapa de Transformação dos dados. Depois de serem selecionados, limpos e pré-processados, os dados são armazenados e formatados de forma apropriada para que os algoritmos possam ser aplicados.



- Chegando na etapa de MD, com a base de dados já selecionada, processada e transformada, foi reunido as partes já manuseadas anteriormente e adicionado à Ferramenta responsável por realizar o carregamento dos dados adequadamente. Após os dados serem carregados na Ferramenta, foi escolhido a técnica encarregada para fazer a tarefa de Mineração. A técnica de árvore de decisão foi o modelo mais apropriado para esse trabalho. O que levou a utilização deste modelo foi dado ao fato da técnica desenvolver um mecanismo de ajuda na tomada de decisões.

RESULTADOS E DISCUSSÕES

Os resultados consistem na MD tendo em base de experimento o SIM, onde foi realizado através da técnica de árvore de decisão e do algoritmo J48, trazendo previsões de ocorrências de óbitos, de modo que as informações extraídas sirvam de interpretação e análise. Com a base de dados do SIM do DATASUS no ano de 2015 com o objetivo de minerar possíveis circunstâncias de ocorrências de óbitos nos Municípios do Estado do Rio Grande do Sul (RS), foi gerada uma árvore com vários resultados. O resultado foi gerado pela técnica de Árvore de Decisão na ferramenta WEKA. Por ser uma técnica descritiva, a árvore gerada descreve toda a base de dados que foi minerada.

Na Figura 1 há uma reconstrução dos resultados a partir da árvore original. A navegação na árvore é feita conforme as informações de entradas. O algoritmo vai buscar esses dados selecionados e irá navegar pela árvore.

Figura 1 - Navegação da Árvore



Na Figura 1 é possível visualizar mais detalhadamente a maneira como funciona a navegação do algoritmo na árvore até chegar no resultado do classificador.

CONSIDERAÇÕES FINAIS

Após o estudo de MD baseada na técnica de Árvore de decisão, foi possível comprovar suas vantagens. O sucesso das árvores de decisão, deve-se ao fato de ser uma técnica



extremamente simples, não necessita de parâmetros de configuração e geralmente tem um bom grau de assertividade. Como se trata de uma ferramenta robusta, seu desempenho consegue ser potente mesmo com uma grande quantidade de dados. A pesquisa em geral permitiu que, através de uma base de dados recolhido do portal do DATASUS, fosse possível prever se uma pessoa está mais propensa a uma determinada circunstância de ocorrência de óbito, considerando o local em que está, o período do ano, sua idade, sexo, etnia, estado civil e escolaridade. O resultado obtido no trabalho realizado servirá como premissa para potencializar previsões mais cautelosas em relação às circunstâncias de óbitos.

REFERÊNCIAS

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. 2009. Technical Report - RT-INF_001-09 Relatório Teórico August, Universidade Federal de Goiás. 2009.

CREPALDI, Paola Guarisso. **Um estudo sobre a árvore de decisão e sua importância na habilidade de aprendizado**. Universidade Católica de Londrina, 2012

DANTAS, Eric Rommel G. JÚNIOR, José Carlos Almeida Patrício. LIMA, Daniel Silva de. AZEVEDO, Ryan Ribeiro de. **O Uso da Descoberta de Conhecimento em Base de Dados para Apoiar a Tomada de Decisões**. V Simpósio de Excelencia em Gestão e Tecnologia, 2008.

AVELAR, Cátia Fabíola Parreira de. **MINERAÇÃO DE DADOS: uma revisão da literatura em Administração**. Universidade Federal de Minas Gerais, 2017.

LORENZETT, Cassio Dal Castel. **Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forest e J48 na tomada de Decisão**. Universidade de Cruz Alta, 2016.

COSTA, Claudio Napolis; COUTINHO, Jonatas Vieira; MAGALHÃES, Lúcia Helena de; ARBEX, Márcio Aarestrup. **Descoberta de Conhecimento em Bases de Dados**. 2005. Graduado em Sistema de Informação pela Faculdade– Ciência da Computação, Universidade Federal de Lavras, 2011.

QUINLAN, J.R. C4.5: **programs for machine learning**. Sydney, Austrália: Morgan Kaufmann Publishers, 1993.

GIASSON, Elvio. **Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado Grande, RS, Brasil**. Universidade Federal do Rio Grande do Sul, 2013.

NORTON, M. J. **Knowledge Discovery in databases**. In: Library Trends, 1999.