



MINERAÇÃO DE DADOS PARA O PERFIL DOS PACIENTES AMBULATORIAIS COM CÂNCER¹

Eduardo Ferreira da Silva², Regis Rodolfo Schuch³, Sandro Sawicki⁴. UNIJUI

Introdução: O conhecimento tornou-se um elemento importante para a obtenção de vantagem competitiva e cresce assim a valorização das informações que podem ser geradas a partir dos bancos de dados das organizações. Antes, a preocupação estava voltada para o armazenamento, visto que a capacidade física dos equipamentos era limitada. Atualmente, o foco está voltado à extração de conhecimento útil a partir da vasta quantidade de dados disponíveis que estão além da capacidade de análise através de técnicas ou métodos manuais. Assim, o problema não reside mais na coleta, processamento ou armazenamento de dados, mas sim na capacidade de análise sobre estes dados, de forma a extrair informações e conhecimento útil que possa ser utilizado pelos gestores das organizações. Frente a isso, as técnicas de Mineração de Dados estão sendo cada vez mais utilizadas pelas organizações e estudadas no meio acadêmico, por oferecerem uma solução adequada na busca do conhecimento implícito nas bases de dados. Este trabalho tem por objetivo explorar a base de dados do Departamento de Informática do SUS, mais especificamente o subsistema de procedimentos ambulatoriais, o Sistema de Informação Ambulatorial, com a finalidade de extrair informações sobre os perfis mais prevalentes dos pacientes que realizam procedimentos no estado do Rio Grande do Sul referentes às categorias de cânceres mais incidentes no Brasil no ano de 2008, segundo o Instituto Nacional do Câncer (INCA). **Materiais e Métodos:** A escolha das neoplasias a serem analisadas baseou-se na publicação de 2007 do INCA, sobre a estimativa de incidência do câncer no Brasil para 2008. Utilizou-se os cânceres de localização primária malignos apresentados como mais incidentes, e a partir daí foi estabelecido um modelo de mineração para cada um. O agrupamento das neoplasias selecionadas baseou-se na Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde – Décima Revisão (CID 10). Na primeira etapa foram definidos os atributos que apresentavam características pertinentes à análise e alguma possível contribuição para a obtenção das regras de associação. Os dados, compreendidos no período de janeiro a dezembro de 2008 e referentes ao RS, foram obtidos através de download no portal do DATASUS. Em seguida foi criada uma base de dados utilizando o Sistema de Gerenciamento de Banco de Dados (SGBD) SQL Server 2005 contendo todas as tabelas que formam o SIA. Após a conclusão dos passos anteriores, foi efetuada a importação dos dados. Após todos os atributos terem sido definidos e passados pelos processos de validação e preparação, partiu-se para a execução do algoritmo de Mineração de Dados. **Resultados:** Após a etapa de Mineração, objetivando um grau de segurança para validar os resultados e na ausência de um parâmetro padrão para validação dos mesmos, foram consideradas interessantes, para o usuário, aquelas regras que satisfaziam parâmetros mínimos de um total de registros igual a 1% e confiabilidade igual a 70%. Esse padrão foi levado em consideração devido ao grande número de patologias presentes em cada grupo de neoplasias, o que implicaria em uma grande ocorrência de regras com baixo valor preditivo. Foi possível identificar o perfil dos pacientes que mais prevaleceram em cada grupo de neoplasias. **Conclusão:** Com os resultados obtidos foi possível identificar os padrões e perfis dos pacientes que realizaram procedimentos ambulatoriais referentes às doenças neoplásicas



CT&I e SOCIEDADE

XVIII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA
XV JORNADA DE PESQUISA
XI JORNADA DE EXTENSÃO

4 a 8 de OUTUBRO de 2010



selecionadas no ano de 2008 no RS. Devem-se fazer algumas ressalvas, visto que a veracidade dos registros é de cunho operacional, ou seja, operador dependente. Vale ainda lembrar que um mesmo paciente pode ter realizado mais de uma vez o mesmo procedimento. Sendo assim, a análise das regras geradas e o objetivo do trabalho é traduzir um padrão e perfil populacional esperado, mas não considerado irrefutável. Apoio: PIBEX/Unijuí.

¹ Projeto de extensão realizado no curso de Sistemas de Informações da Unijuí.

² Bolsista PIBIC, aluno do curso de Ciência da Computação, da UNIJUÍ.

³ Bolsista PIBEX, Aluno do curso Sistemas de Informações, da UNIJUÍ.

⁴ Professor Orientador.