

DESCOBERTA DE PADRÕES EM ORDENS DE SERVIÇO DE TECNOLOGIA DA INFORMAÇÃO EM HOSPITAL

RESUMO

Este artigo apresenta um estudo de caso de aplicação do KDD (*Knowledge Discovery in Databases*) sobre Ordens de Serviço (OS) de informática de uma instituição hospitalar do Paraná, objetivando identificar novas práticas a serem aplicadas na gestão. Esta pesquisa surgiu da necessidade de conhecer os padrões de relacionamento entre os diversos setores do hospital com a área de informática, no que se refere à solicitação de atendimentos a problemas de Tecnologia de Informação (TI). Dessa forma, seria possível avaliar a atuação efetiva do setor, bem como identificar eventuais desvios ou não conformidades. Para os experimentos foi utilizada a tarefa de classificação, sendo avaliados três métodos, possibilitando a obtenção de regras com potencial de validação por especialista. O resultado produzido neste estudo demonstrou a aplicabilidade do KDD na obtenção de conhecimento relevante para auxílio à tomada de decisão com vistas ao investimento em melhores práticas para ganho de qualidade dos serviços de TI. Dentre eles, destacam-se: o desenvolvimento de planos de ação para ajustes de inconsistências no uso da OS pelos setores para demanda de ações de TI; o estabelecimento de novos parâmetros de priorização dos atendimentos emergenciais dos setores demandantes; o ajuste, implementação e divulgação de novos de procedimentos de atendimento das OS; e em consequência o ganho de agilidade e qualidade no atendimento e a melhora no relacionamento da TI com os demais setores.

Palavras-chaves: mineração de dados, gestão hospitalar, governança de tecnologia de informação.

PATTERN DISCOVERY IN IT SERVICE ORDERS FOR HOSPITALS

ABSTRACT

The article presents a case study about KDD (*Knowledge Discovery in Databases*) application in IT (information technology) Service Orders (OS) in a hospital in the State of Paraná, aiming to identify new practices to be applied in IT management. The research arose from the need to identify relationship patterns between hospital departments and IT, based on their IT problem demands. From these patterns it could be possible to assess the effective performance of the sector and identify any deviations or non-conformities. The experiment used classification and evaluated three methods, which made it possible to obtain rules capable of expert validation. The results produced in this study demonstrated the KDD applicability to obtain relevant knowledge to aid decision-making for investment in the betterment of IT services. Among them: the development of action plans for adjustment of inconsistencies in IT-managed OS requests; the establishment of new prioritization standards for emergency requests; the improvement, implementation and dissemination of new OS work procedures; and, as a consequence, a increase in agility and service quality and an improvement th IT relationship with other hospital sectors.

Keywords: data mining, hospital management, IT management

Gilson Eloy Ferndandes França¹
Deborah Ribeiro Carvalho²
Denise Fukumi Tsunoda³

¹ Mestre pela Universidade Tecnológica do Paraná – UTFPR, Brasil
Sócio diretor Conseiller Tecnologia da Informação e Conhecimento
E-mail: gfranca@weega.com.br

² Doutor pela Universidade Federal do Rio de Janeiro – UFRJ, Brasil
Professor pela Pontifícia Universidade Católica do Paraná – PUCPR, Brasil
E-mail: ribeiro.carvalho@pucpr.br

³ Doutora pela Universidade Tecnológica do Paraná - UTFPR, Brasil
Professora pela Universidade Federal do Paraná - UFPR, Brasil
E-mail: dtsunoda@ufpr.br

1 INTRODUÇÃO

Os hospitais de todo o país estão passando por transformações com o crescimento do acesso da população a novas especialidades de atendimento médico, tecnologias de diagnóstico e de tratamento. Tal diversidade e complexidade exige cada vez mais da capacidade de gestão envolvendo um grande número de atividades operacionais e de apoio, necessárias para manter toda a estrutura hospitalar funcionando eficientemente.

Desde o final do século XX, a aplicação da tecnologia da informação (TI) no ambiente hospitalar tem avançando sobremaneira, produzindo efeitos benéficos na gestão e, principalmente, no tratamento e acompanhamento das necessidades de saúde dos pacientes, que são o elemento central das instituições de saúde (Salu, 2013).

Dentre os principais recursos e ferramentas de desenvolvimento, monitoramento e controle operacional de um hospital está o sistema de ordens de serviço (OS), geralmente automatizadas nos sistemas ERP (*Enterprise Resource Planning*), utilizado para requisição e acompanhamento de atendimento especializado pelos setores.

Todas as OS demandadas pelos demais setores ao setor de TI são armazenadas em Bases de Dados (BD) do sistema ERP, e incorporam um grande volume de informações sobre os incidentes e problemas tratados, bem como sobre os seus respectivos padrões de ocorrência e de atendimento. Ocorre que, dado o seu tamanho significativo, em quantidade de registros e atributos, tais informações dificilmente conseguem ser analisadas de forma efetiva por meio dos relatórios tradicionais gerados pelo sistema ERP, dificultando, dessa forma, a gestão do serviço.

Outro fator crítico, identificado por Malagón-Londoño, Morera e Laverde (2010), é que os gestores dos hospitais têm dificuldade de utilizar as informações processadas pela TI na tomada de decisão, pois estas não atendem a seus requisitos de compreensibilidade e adequação às suas necessidades.

Diante de tudo isso, a metodologia KDD (*Knowledge Discovery in Databases*) foi considerada uma alternativa para obtenção de novos conhecimentos, mais complexos e profundos, que permitissem avaliar, sob aspectos diversificados, os principais e mais relevantes problemas de TI existentes, bem como o nível de relacionamento da equipe de TI com os demais setores do hospital. Os resultados advindos deste experimento devem auxiliar na identificação da necessidade de implementação de

ajustes e melhores práticas na TI, com objetivo de ganho de qualidade nos serviços de TI.

Os resultados produzidos pela aplicação do KDD demonstraram a possibilidade de utilização dessa metodologia para obtenção de novos conhecimentos capazes de permitir ao gestor de TI um melhor direcionamento de seus esforços organizacionais.

Para a instituição objeto deste estudo de caso, foi possível perceber alguns indicativos de inconsistências operacionais no atendimento das necessidades de TI dos setores, os quais demandaram a implantação de planos de ação para ajustar tais dissonâncias.

Como contribuição, este artigo destaca a viabilidade da aplicação da mineração de dados sobre grandes volumes gerados nas bases de OS de TI das instituições médicas, constituindo fontes importantes de pesquisa para compreensão do relacionamento do referido setor com toda a organização, possibilitando o desenvolvimento de ações de planejamento e organização que permitam torná-lo cada vez mais efetivo no suporte às diretrizes estratégicas e operacionais.

2 A IMPORTÂNCIA DA TECNOLOGIA DA INFORMAÇÃO PARA GESTÃO HOSPITALAR

Hospitais são instituições que compõem parte de um sistema coordenado de saúde, com função de dispensar assistência médica, preventiva e curativa, desenvolvimento de atividades de natureza comunitária, pesquisa e formação profissional (Malferrari et al., 1983).

São destinados ao diagnóstico e tratamento de pessoas segundo um conjunto de especialidades médicas, podendo também servir para pesquisa e ensino. Geralmente estão divididos em: uma área assistencial, para o atendimento dos doentes; outra administrativa/contábil, responsável pela operação das unidades, pelo controle financeiro-contábil e humano; e por fim uma área responsável pelo planejamento e gestão de toda a instituição. A estrutura hierárquica dos hospitais varia conforme sua classificação, em função do nível de atenção em saúde, porte e tipo de gestão (pública ou privada) (Malagón-Londoño, Morera, & Laverde, 2010).

Segundo Lima Gonçalves (1983, como citado em Hauser, 2006, p. 3), as principais funções de um hospital são: prestação de atendimento médico e complementar aos doentes em regime de internação; desenvolvimento, sempre que possível, de atividades

de natureza preventiva; participação em programas de natureza comunitária, procurando atingir o contexto sociofamiliar dos doentes, incluindo aqui a educação em saúde, que abrange a divulgação dos conceitos de promoção, proteção e prevenção da saúde; integração ativa no sistema de saúde.

Os hospitais estão entre as estruturas organizacionais mais complexas da sociedade moderna, pois incorporam uma grande variedade de funções, cada vez mais sofisticadas e especializadas, para recuperar, manter e incrementar os padrões de saúde das pessoas. São instituições que, em geral, precisam operar com toda a sua estrutura e as suas especialidades durante todos os dias do ano vinte e quatro horas por dia (Senhoras, 2007).

Tudo isso demanda grande quantidade de processos que transacionam cada vez maiores volumes de dados e informações entre as diversas unidades componentes da instituição médica, gerando uma interdependência crescente e vital. A operacionalização fluida de tamanho volume de informações só é possível com a utilização de recursos tecnológicos de gestão de informações. Segundo Pinochet (2011), a TI hospitalar já ultrapassou o modelo padrão de processamento de dados que ficava restrito às áreas administrativas das organizações. Agora ela atua de forma essencial desde a gestão das atividades assistenciais, no diagnóstico até o atendimento do paciente. Sem isso, as instituições poderiam ser colocadas em risco pela crescente demanda de atendimentos médicos com alta complexidade e custos operacionais consideravelmente maiores.

Pereira *et al.* (2012) colocam que a tecnologia da informação, no setor da saúde, por meio de seus sistemas informatizados, auxilia os hospitais no cumprimento de seus objetivos e funções pelo suporte às atividades assistenciais, administrativas e gerenciais.

Segundo Gutierrez (2011), a utilização da TI tem o potencial de produzir ganhos de eficiência e redução de falhas no processo assistencial dos hospitais com importante melhoria da qualidade dos serviços, com custos mais otimizados.

Hannah, Ball e Edwards (2009, como citado em Pinochet, 2011, p. 385) consideram que os sistemas de informação para saúde devem ser classificados em três tipos: sistemas *stand alone*: sistemas limitados quanto ao objetivo e a abrangência, geralmente de uso específico. Exemplos destes seriam: os sistemas incluídos nesse grupo são direcionados a laboratórios, controle financeiro, radiologia, eletrocardiografia, controle de funções pulmonares, sistema de farmácia e nutrição. O segundo tipo é composto de sistema de informação hospitalar, que geralmente está baseado em uma rede, e é composto por algumas funções clínicas e administrativo-financeiras. Por último, sistemas

corporativos de informação em saúde, chamados de *Enterprise Resource Planning* (ERP) ou Sistema Integrado de Gestão Empresarial (SIGE ou SIG). Segundo Campanairio, Maccari, e Paulo (2004), os ERP democratizam as informações para todas as áreas das organizações facilitando a tomada de decisão, com ganho de produtividade, redução de custos e eliminação de problemas de comunicação entre sistemas diferentes. Nos hospitais, abrangem todas as funções em três níveis: assistencial, administrativo e de gestão, de forma integrada. Eles estão associados a diversos tipos de recursos para diagnósticos, em que os registros são capturados e depositados em vários tipos de mídia, incluindo som, imagem, animação e impressão. Todas as informações, nos diversos formatos, são armazenadas de forma centralizada em bancos de dados, utilizando a abordagem dos *Data Warehouse*⁴ (DW).

Burmester (2012) comenta que um estudo recente do Congresso dos EUA, sobre a reforma do sistema de saúde americano, recomenda a aplicação de técnicas de gestão tradicionalmente de outros segmentos, e entre os elementos importantes a serem contemplados estavam: a condução dos hospitais a uma administração estratégica; utilização da informação como elemento essencial à análise dos resultados e subsídios ao planejamento; e gestão dos processos de atendimento do hospital.

As organizações hospitalares estão cada vez mais dependentes dessas tecnologias como fator de gerenciamento eficiente dos serviços produzindo ganhos de competitividade e qualidade no atendimento do paciente. Pinochet (2011) destaca que todos os avanços tecnológicos disponíveis, em especial em TI, nos mais diversos setores, exigem um alto investimento, têm um custo operacional e de manutenção considerável e, em alguns casos, eles se tornam obsoletos rapidamente. Portanto, a efetiva gestão desses recursos, alinhada às questões estratégicas da organização, é fundamental. Isso depende de se ter as informações corretas à disposição no momento correto, no formato acessível para que possa tomar decisões assertivas.

3 GOVERNANÇA DE TI

O sucesso das organizações passa essencialmente pela gestão de operações, de transações e do conhecimento, de forma a obter sustentabilidade econômica e social para as suas atividades no mercado onde atuam. Para tanto, a grande maioria das organizações investe em TI como suporte para a manutenção e crescimento do seu negócio. Ainda assim, implementar uma TI capaz de efetivamente entregar tais benefícios exige a consciência dos riscos e

⁴ Local de armazenamento consolidado de informações de uma organização.

desafios envolvidos.

O setor de TI das empresas é o responsável pelo planejamento, implementação, uso e manutenção dos recursos de infraestrutura e ferramentas de informática dentro das organizações. Isso significa fazê-lo buscando ganho permanente de desempenho com eficiência econômica (Becker & Knackstedt, 2009). É uma área capaz de atuar de forma multidisciplinar por toda a empresa. São diversas as contribuições que a TI pode promover nos mais variados campos, em todos os níveis da organização, utilizando recursos tecnológicos altamente especializados, demandando investimentos consideráveis e profissionais qualificados. Tudo sempre como provedora do suporte fundamental à conquista dos objetivos estratégicos da organização.

Dada a grande importância e os elevados riscos que a TI impõe às organizações, desenvolveu-se o conceito da governança de TI (GTI), que é derivado da governança corporativa. A GTI tem como objetivos: gerar maiores garantias, maior transparência e controle sobre os investimentos, recursos e no desenvolvimento das ações da TI; estabelecer e manter o alinhamento da TI com os objetivos estratégicos corporativos; e buscar a excelência operacional da TI.

Segundo Robinson (2007), estabelecer um programa de governança de TI representa a combinação de análise de processos sistemáticos e aspectos da ciência comportamental, uma vez que a governança de TI permeia toda a organização nos seus diversos níveis gerenciais e que cada organização possui personalidade e características únicas, o que, em princípio, exige um trabalho personalizado para cada modelo organizacional encontrado.

Segundo o IT Governance Institute (ITGI), a aplicação efetiva da governança de TI é responsabilidade da direção e da gerência da organização. A governança de TI é uma parte integrante da governança corporativa e consiste nas estruturas e nos processos de liderança e organizacionais que assegurem que a TI suporte e fortalece as estratégias e objetivos da organização.

A utilização de metodologia e melhores práticas pode ser elemento crítico para garantir o controle e a governança da TI. Algumas das metodologias e melhores práticas são: CobiT, ITIL, PMBok, entre outros. O *Control Objectives for Information and related Technology* (CobiT) é um guia para gestão de TI desenvolvido pelo Information Systems Audit and Control Foundation (ISACF), orientado ao negócio para o auxílio de três públicos de forma distinta: gerentes, usuários e auditores. O *Information Technology Infrastructure Library* (ITIL) foi desenvolvido pela secretaria de comércio do Governo Inglês (Office of Government Commerce – OGC) e é considerado uma biblioteca de melhores práticas para operação e gestão de serviços de infraestrutura de TI. O *Project Management Body of*

Knowledge (PMBok), desenvolvido pelo Project Management Institute (PMI) é um conjunto de melhores práticas para gestão de projetos.

A utilização das melhores práticas permite às organizações alcançarem uma maior efetividade no alcance dos seus objetivos estratégicos e controle do risco pela adoção da TI em seus processos.

As organizações com elevada maturidade de governança de TI têm desempenho significativamente superior, pois estão focadas principalmente nas necessidades dos clientes e do negócio. Conseqüentemente, a governança passa a ser elemento gerador de valor e de diferenciação para o mercado (Bordoloi, 2012).

4 KDD E MINERAÇÃO DE DADOS

Com o crescimento do uso da TI, grandes volumes de informações passaram a ser produzidos nas organizações, estimulados pelo aumento da capacidade e redução dos custos de processamento e armazenamento de dados, bem como pelo surgimento de novos formatos e tecnologias de compartilhamento e pela rápida evolução da internet (Maimon & Rokach, 2010; Witten, Frank, & Hall, 2011). Para Rezende (2003), a principal motivação para esse crescimento está relacionada à facilidade para coleta, armazenamento e gerenciamento de grande quantidade de dados.

Todo esse volume de dados pode conter informações importantes capazes de auxiliar a tomada de decisão, gerando maior efetividade mercadológica (Larose, 2005). Mas, para tanto, é essencial ser capaz de extrair informação ou conhecimento útil de tal quantidade de dados armazenados.

Pode-se então pensar na utilização das ferramentas tradicionais de TI para busca desse conhecimento. Mas estas são capazes apenas de exibir os dados formatados, pois foram concebidas para gerar relatórios simplificados de atividades como: listar itens vendidos, verificar estoques de vendas, desenhar gráficos etc., sem produzir qualquer tipo inferência (Rezende, 2003), tornando a capacidade de tomada de decisão dos gestores limitada.

Entre as técnicas não tradicionais destaca-se o processo *Knowledge-Discovery in Databases* (KDD). Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), o processo de extração de conhecimento de bases de dados tem o objetivo de encontrar conhecimento a partir de um conjunto de dados para ser utilizado em um processo decisório. Um requisito importante é que o conhecimento descoberto seja compreensível, além de útil e de interesse para os usuários, ou seja, tomadores de decisões.

Segundo Goldschmidt e Passos (2005), o KDD é composto por três componentes básicos: o problema, os recursos para solução do problema e os resultados obtidos pela utilização dos recursos para se

encontrar a solução do problema. O problema requer um conjunto dos dados que serão utilizados para aplicação do processo, um ou um grupo de especialistas sobre o tema ao qual se busca conhecimento e objetivos claros quanto ao que se pretende obter através dos dados. Os recursos representam o profissional capaz de conduzir e executar o processo do KDD, interagindo com o especialista e os recursos computacionais a serem utilizados em todo o processo, podendo ser: software (ferramenta de KDD) ou hardware (plataforma computacional). Os resultados obtidos seriam os modelos de conhecimento obtidos pela aplicação do processo.

O KDD segue cinco etapas bem definidas que ocorrem de forma geralmente sequencial não definitiva, uma vez que em qualquer delas pode haver a necessidade de retorno à etapa anterior para refinamentos importantes para ganho de eficiência das etapas subsequentes.

Todo o processo inicia com a definição dos objetivos a serem alcançados com a aplicação do método e a preparação dos recursos necessários para sua execução, incluindo: dados, conhecimentos sobre o ambiente que se deseja trabalhar, profissionais e recursos necessários para o experimento. Para Fayyad, Piatetsky-Shapiro e Smyth (1996), o conhecimento a ser buscado deve estar orientado sempre às necessidades dos usuários.

Em seguida é feita a seleção e preparação dos dados, orientada geralmente por especialista. Esta requer um trabalho cuidadoso de seleção e extração dos dados, a partir de grandes conjuntos de dados (bases de dados), produzindo um novo conjunto (DW) em que são realizados refinamento e complementação conforme os objetivos estabelecidos no início do método. A etapa de mineração pode ser mais efetiva quanto maior a qualidade resultante do trabalho no pré-processamento. Goldschmidt e Passos (2005) dizem que o pré-processamento tem como objetivo preparar os dados obtidos para etapa de mineração. Para tanto, ela envolve todo o processo de obtenção, organização e tratamento dos dados.

Por fim, há a transformação dos dados selecionados e preparados para o formato apropriado à execução da ferramenta de mineração de dados selecionada como mais adequada ao tipo e formato de dados disponibilizados.

A mineração de dados (MD), quarta etapa do KDD, é o processo de descoberta de novas correlações, padrões e tendências significativos, obtidos pelo uso de tecnologias de reconhecimento de padrões, estatística e matemática, a partir de grandes volumes de dados armazenados em bancos de dados ou Data Warehouse (DW) (Larose, 2005). Para Fayyad, Piatetsky-Shapiro e Smyth (1996), MD é definida como “extração de conhecimento de base de dados e o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos

dados”.

Segundo Galvão e Marin (2009), uma parte fundamental da MD é a definição do(s) algoritmo(s) que será(ão) utilizado(s) para obtenção de resultados que atendam aos objetivos do estudo. As tarefas possíveis de um algoritmo de extração de padrões podem ser agrupadas em atividades preditivas e descritivas. As tarefas de predição são: classificação e regressão. As principais tarefas descritivas são: agrupamento, associação e sumarização.

Uma das técnicas de mineração de dados mais conhecidas e utilizadas é a classificação, que busca a descoberta de um modelo preditivo, ou seja, funções que permitam o mapeamento de dados em um conjunto de classes que possa ser utilizado para classificar novos conjuntos de dados desconhecidos. Segundo Han e Kamber (2006), a classificação é o processo de busca de um modelo ou função que descreva ou distinga classes de dados que possam ser utilizados para predição de classes de objetos desconhecidos. Sua representação pode ocorrer de diversas formas, como: regras de classificação (Se – Então), árvores de decisão, fórmulas matemáticas ou redes neurais. Nesta pesquisa, trabalhou-se com regras de classificação e árvores de decisão.

As árvores de decisão são estruturas gráficas organizadas em nós, ramos e folhas. Cada nó representa um teste em um valor de atributo, os ramos representam, cada um, um resultado do teste, e as folhas, classes ou distribuições de classe. As árvores de decisão geralmente podem ser transformadas em regras de classificação. Para Larose (2005), a conversão do modelo gráfico das árvores em regras de decisão dá ao modelo um dos seus mais relevantes aspectos: a conversibilidade em regras (se antecedente, então consequente), o que permite melhor interpretação do modelo de decisão. Algumas das aplicações mais comuns de classificação são: aprovação de crédito, marketing direto, detecção de fraudes, entre outras. No QUADRO 01 estão apresentadas as principais características dos três algoritmos de classificação utilizados nesta pesquisa.

A definição do algoritmo, segundo Resende (2005), deve estar atrelada aos objetivos definidos pela MD e também aos padrões de representação de dados utilizados para busca de conhecimento. A qualidade da solução encontrada está diretamente relacionada com o conteúdo levado ao algoritmo. Por isso as etapas antecedentes ao processo de mineração devem ser tão exaustivamente trabalhadas, permitindo maior grau de assertividade nas suas etapas consequentes.

A extração de conhecimento de uma base de dados utiliza ferramentas de TI que podem implementar um ou mais algoritmos. Essas ferramentas simplificam e agilizam sobremaneira esse processo, pois geralmente são intuitivas e possuem uma interface amigável. Existem no mercado diversas ferramentas de uso comercial, acadêmico e científico, desenvolvidas por

universidades e laboratórios de pesquisa de diversas partes do mundo, sendo as principais, segundo Thuraisingham (1999, como citado em Resende, 2005, p. 328-9): PolyAnalyst; Magnum Opus; XpertRule Miner; DataMite; Microsoft Data Analyzer 2002; Oracle 9i Data Mining; Darwin; MineSet; WEKA; Intelligent Miner; MCC++; See5; Cubist; Data-Miner Software Kit.

A última etapa do KDD, o pós-processamento, avalia os resultados obtidos pelo foco a qualidade da informação produzida. Qualidade que pode ser traduzida por fatores como a capacidade de se gerar

informação evidente, inovadora, útil e aplicável sob o ponto de vista da especialidade e do problema que se desejou trabalhar. Goldschmidt e Passos (2005) colocam como atividades do pós-processamento: a visualização, análise e interpretação, pelo especialista, do modelo de conhecimento produzido na MD. É muito importante que os modelos resultantes da mineração sejam avaliados em relação à qualidade e eficácia para que possam ser aplicados a situações práticas, uma vez que todo o trabalho da MD demanda um investimento considerável da empresa (Larose, 2005).

QUADRO 01 – Principais características dos algoritmos classificadores selecionados

Itens	C 4.5 / J 48	JRIP	PART
Classe do algoritmo	Algoritmo preditivo	Algoritmo preditivo	Algoritmo preditivo
Formato	Árvore de decisão	Regras de decisão	Árvores de decisão parcial
Operação do algoritmo	<p>Particiona recursivamente um conjunto de treinamento até que cada subconjunto obtido deste particionamento contenha casos de uma única classe.</p> <p>Aplica o critério de divisão e conquista na construção da árvore, e o critério guloso para escolha de melhores partições e melhores atributos. Utiliza a entropia como fator de seleção das melhores características para cada divisão.</p>	<p>Opera em dois estágios: primeiro descobre um conjunto de nós iniciais (estratégia de dividir e conquistar) e depois refina ajustando aqueles individuais.</p> <p>Esta estratégia determina o nó que melhor caracteriza a segmentação dos conjuntos. Esse procedimento é repetido para cada um dos subconjuntos.</p> <p>Utiliza técnicas de heurísticas.</p>	<p>Não precisa gerar uma otimização para produzir um conjunto de regras precisas (sua maior vantagem). Utiliza a estratégia de dividir e conquistar, simultaneamente constrói a regra, remove as instâncias cobertas recursivamente até não existirem mais regras.</p> <p>Este modelo a torna mais ágil e flexível, uma vez que não precisa construir toda a árvore para gerar uma única regra.</p>
Objetivo do processo de otimização global	O conjunto inicial de regras geradas é muito grande e redundante, pois é gerado a partir de uma árvore de decisão.	Aumentar a precisão do conjunto de regras, substituindo ou revisando regras individuais.	Não precisa executar

Fonte: Frank e Witten (1998) Barbosa, Carneiro e Tavares [s.d.]

5 ENCAMINHAMENTOS METODOLÓGICOS

Para esta pesquisa foi selecionada uma instituição hospitalar do Estado do Paraná que atendesse a alguns requisitos: possuísse ferramenta ERP, contasse com uma equipe de TI para desenvolvimento e manutenção e que utilizasse a ferramenta de OS para receber, gerenciar e retornar as solicitações de serviços. A instituição médica

selecionada foi um hospital que será nomeado Hospital H para preservar a identidade do mesmo, uma vez que não influenciará de qualquer forma os resultados.

Foram utilizados como ferramentas para levantamento, manuseio e execução da pesquisa a linguagem *Structured Query Language* (SQL), a planilha eletrônica MS – Excel e o software de domínio público de KDD denominado *Waikato Environment for Knowledge Analysis* (Weka) (Witten, Frank, & Hall,

2011).

A metodologia que estruturou o trabalho foi o KDD (*Knowledge Discovery in Databases*), por meio de suas cinco fases, da seleção ao pós-processamento, conforme detalhado.

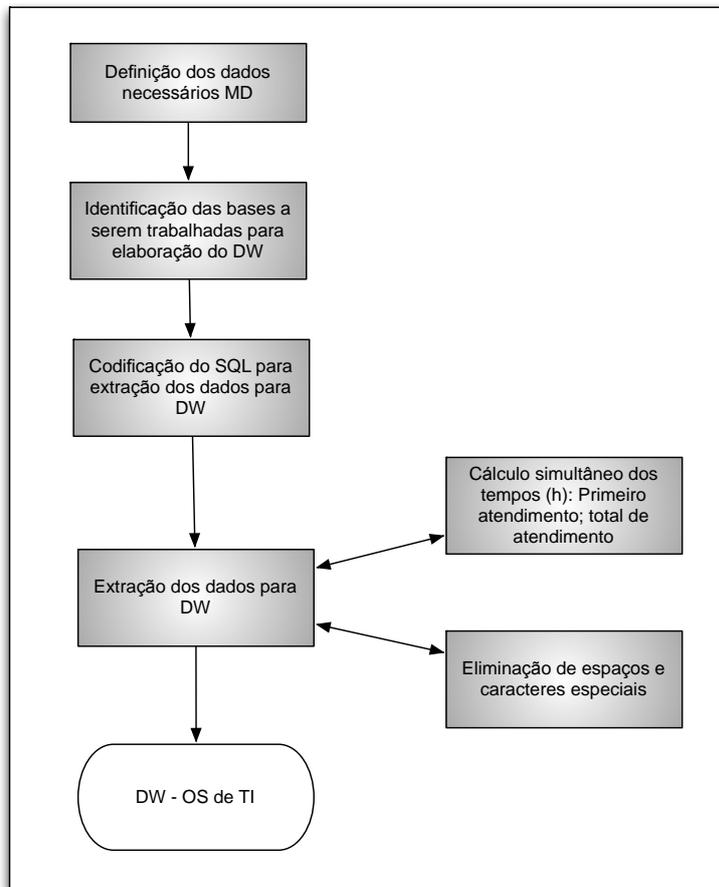
1) Seleção:

Identificação e seleção de uma instituição hospitalar que atendesse a condições já relacionadas de uso de ferramenta ERP, banco de dados e das OS como instrumento de gerenciamento dos atendimentos das necessidades dos demais setores e estabelecimentos do hospital pela equipe de TI. Nessa etapa buscou-se compreender o modelo operacional do sistema de

solicitações, bem como todo o ambiente vivenciado pela instituição em relação ao atendimento das necessidades pela TI. Tudo isso só foi possível com o acompanhamento de um especialista da área, que auxiliou na evidenciação dos pontos relevantes e críticos do processo em questão.

O processo de elaboração do *data warehouse* (DW) baseou-se nas conclusões obtidas do levantamento do ambiente e seu escopo operacional, seguindo os passos detalhados na Figura 1. A descrição dos atributos dos dados extraídos das bases de dados através desse processo e que foram utilizados no pré-processamento estão detalhados no QUADRO 2.

Figura 1: Processo de extração de dados para DW



QUADRO 2 – Relação dos atributos disponíveis no DW

Atributo	Descrição	Tipo
OS	Número da OS	Numérico
SOLICITANTE	Nome do solicitante	Nominal
SETOR	Setor solicitante	Nominal
GRUPO_TRAB	Código do grupo de trab. da TI	Numérico
EQUIPAMENTO	Número do equipamento origem da OS	Numérico
LOCAL	Localização do equipamento da OS	Nominal
PRIOR	Prioridade da OS	Nominal
CAUSA_DANO	Descrição do dano	Nominal
TIPO_SOLUCAO	Descrição da solução	Nominal

DT_AB_OS	Data de abertura da OS	Numérico
ANALISTA_TI	Nome do analista de TI que atendeu a OS	Nominal
INICIO_REAL	Data de início do atendimento da OS	Numérico
FIM_REAL	Data de conclusão do atendimento da OS	Numérico
TMP_INI_AT	Tempo para o início do atendimento	Numérico
TMP_EX_OS	Tempo para o atendimento	Numérico
TMP_TT_OS	Tempo total da solicitação à conclusão	Numérico
STATUS_OS	Situação da OS	Numérico
ESTABELECIMENTO	Estabelecimento solicitante	Nominal

2) Pré-processamento:

A preparação do conteúdo do DW para utilização dos métodos de extração de conhecimento na ferramenta selecionada WEKA exigiu diversos ajustes para produzir maior eficiência possível na execução dos algoritmos de mineração de dados. Para tanto, os atributos ou instâncias foram ajustados, transformados ou até mesmo eliminados. Ajustados ou transformados no sentido de padronizá-los em forma e, quando necessário, conteúdo. Por exemplo: algumas nomenclaturas utilizadas para descrever um mesmo atributo ocorriam de diversas formas. Diversos atributos continham caracteres não aceitos pela ferramenta de mineração e precisaram ser alterados ou eliminados. Em outras circunstâncias faltavam conteúdos de determinados atributos fundamentais. Por vezes o conteúdo de grande parte dos atributos de uma instância estava todo ou quase todo ilegível, determinando que fosse excluído para não prejudicar a execução da ferramenta. Alguns algoritmos só trabalham com números reais, outros, só com valores nominais. Em suma, nessa etapa todos os ajustes necessários foram realizados para garantir que o WEKA pudesse operar com todos os dados do DW, produzindo os resultados desejados.

3) Mineração:

Dadas as características dos dados (não estavam estruturados, havia atributos numéricos e nominais, havia possibilidade de identificação de mais de um atributo meta para mineração, havia a possibilidade de geração de regras de decisão a partir de classificação) obtidos nas bases de dados de OS, decidiu-se utilizar a técnica de classificação e os algoritmos (no padrão WEKA): J48, PART, JRIP.

A execução de algoritmos de predição: PART e RJ48 com ajustes nos parâmetros de confiança e atributos meta demonstrou que os resultados não geravam informação interessante e conclusiva para atendimento dos objetivos da pesquisa.

A partir dessas conclusões da primeira mineração, decidiu-se então trabalhar novamente sobre o DW, eliminando atributos e conteúdos considerados irrelevantes, incluindo novo atributo COMPLEX (complexidade) e normalizando o atributo DANO_BREVE. Juntamente com um especialista do ERP Hospitalar, foram avaliadas todas as ocorrências do atributo DANO_BREVE e classificadas em grupos,

reduzindo assim o número de ocorrências do mesmo. A partir dos atributos CAUSA_DANO, TIPO_SOLUCAO e DANO_BREVE, pode-se estabelecer valores para o novo atributo COMPLEX que representa o nível de dificuldade técnica para solução do problema elencado na respectiva OS. COMPLEX tem em seu escopo: A = alta complexidade; M = Média complexidade; B = Baixa complexidade. Foram eliminados os atributos: OS, SOLICITANTE, GRUPOT, DT_AB_OS, FIM_REAL, INICIO_REAL, pois estes não contribuíam no processo. Da mesma forma, CAUSA_DANO e TIPO_SOLUCAO, por não haver uma normalização adequada e também por haver cerca de 60% de seu conteúdo vazio, o que não produzia resultado efetivo na mineração de dados.

Para melhor avaliação comparativa dos resultados obtidos pela execução dos três algoritmos de classificação, utilizou-se a ferramenta WEKA, aplicando-os com o mesmo modo, com os mesmos atributos e testando-os com quatro atributos metas (ANALISTA_TI, SETOR, DANO_BREVE e COMPLEX) cada um, conforme apresentado no QUADRO 02.

A comparação dos resultados das execuções permitiu uma avaliação em relação a: precisão, compreensibilidade e utilidade do conhecimento gerado.

6 RESULTADOS E PÓS-PROCESSAMENTO

Identificar os padrões existentes no atendimento das necessidades dos setores do hospital, bem como os problemas de TI mais relevantes demandados, utilizando um método capaz de trabalhar com grandes volumes de dados, para estabelecer novos padrões de qualidade nos serviços de TI, indicou a necessidade desta pesquisa executada.

Os resultados advindos do experimento que foi orientado nas quatro primeiras fases do método, aqui culmina na avaliação, análise e interpretação dos resultados da mineração de dados, de forma a permitir identificar a necessidade de implementação de melhores práticas na TI, com objetivo de ganho de qualidade nos serviços de TI.

Durante as três séries de minerações executadas utilizando cada um dos três algoritmos de

classificação, avaliaram-se dados que pudessem refletir o nível de qualidade da informação produzida, tais como: a proporção entre as variáveis que, segundo a ferramenta, foram classificadas corretamente em relação às que não foram classificadas corretamente; o nível de cobertura que o algoritmo conseguiu sobre as ocorrências disponíveis no DW trabalhado, e a produção de conhecimento útil e interessante. Segundo Larose (2005), a busca de conhecimento pela MD não deve se ater somente a um único método, mas sim na confluência dos resultados obtidos em um conjunto de modelos de MD.

No tocante à classificação correta das variáveis, os resultados obtidos foram satisfatórios para 05 das 12 minerações realizadas, ou seja, sempre próximos ou pouco superiores a 60%, ótimos para 05 das 12 minerações que tiveram mais de 80% de classificações corretas e, por fim, bons para somente 02 das minerações ficaram entre 70% e 80%.

Para cobertura, ou seja, número de instâncias utilizadas para identificação de padrões, todos os resultados foram superiores a 95%, o que representa uma grande abrangência para o conhecimento produzido. Ainda assim, somente esses fatores não garantem a qualidade do conhecimento gerado. Há que

se avaliar a sua compreensibilidade e utilidade, mesmo que seja necessário um refinamento.

A classificação com o J48, apesar de gerar maiores proporções de variáveis corretamente classificadas e matrizes de confusão mais equilibradas, também gerou regras com menor índice de erro, não conseguiu produzir efetivamente regras de decisão que atendessem aos requisitos de qualidade de informação. No tocante à estrutura das árvores de decisão criadas, estas apresentaram muitos ramos e nós (entre 150 e 650), o que torna a sua interpretação difícil. Além disso, grande parte das informações contidas nestes eram óbvias (relacionavam apenas o setor e o respectivo equipamento atendido) e portando deveriam ser descatadas. Somente na mineração executada com atributo meta ANALISTA_TI é que se obteve algum conhecimento útil (conforme exemplo apresentado abaixo), ainda assim em uma proporção baixa em relação a todo o conjunto. Todas as execuções trabalharam com índices de cobertura altos, sempre acima de 90%. Os resultados portanto não atenderam às expectativas e objetivos da pesquisa, provavelmente pelas suas características de funcionamento do algoritmo em relação ao nível de ruído existente na base.

Exemplo de parte de uma árvore de decisão gerada na MD do atributo ANALISTA_TI:

```
GRUPO_TRAB <= 1
| TMP_TT_OS <= 2
| | DANO_BREVE = Acesso_uso_rede_Internet: DS (2.0)
| | DANO_BREVE = Problemas_ERP
| | | COMPLEX = M
| | | | TMP_EX_OS <= 0: LA (8.18/3.18)
| | | | TMP_EX_OS > 0: CA (10.22/2.22)
| | | COMPLEX = B: DS (20.0/10.0)
| | | COMPLEX = A: DS (4.0)
```

Os algoritmos geradores de regras de decisão, dadas as suas características mais agressivas e efetivas de poda, provavelmente conseguiram atuar sobre a base produzindo regras de maior interesse, mesmo que com proporções de erro de classificação (erro de classificação de variáveis) maiores que as do J48. Ainda assim, todos mantiveram um índice de cobertura superior a 90%, o que representa o nível de utilização das instâncias do DW. Dentre os dois, o JRIP se

mostrou mais eficiente, apesar de gerar menos regras, ele produziu conhecimento mais interessante que o PART. As execuções de mineração que utilizaram os atributos ANALISTA_TI e DANO_BREVE foram as que regras mais interessantes e úteis geraram, segundo a avaliação do especialista.

Exemplo de regra de decisão gerada pelos algoritmos:

```
(GRUPO_TRAB >= 11) and (TMP_EX_OS >= 1) and (COMPLEX = B) =>
DANO_BREVE=Problemas_Est_Trab (265.0/100.0)
```

```
TMP_EX_OS <= 51 AND EQUIP = SIS_ERP AND DANO_BREVE = ERP_Prescricao AND
ANALISTA_TI = AT: B (10.56/2.56)
```

```
TMP_EX_OS <= 53 AND DANO_BREVE = Problemas_Perifericos AND
EQUIP = BrotherHL5250DN: B (37.0/2.0)
```

A primeira regra demonstra que a equipe de TI 11 atende problemas em estações de trabalho, com

nível de complexidade baixa, com tempo mínimo de 1 minuto. A segunda regra estabelece que o analista AT

atende a problemas de prescrição, de baixa complexidade, com tempo médio inferior a 51 minutos.

Regras como essas são passíveis de transformação para serem utilizadas no ambiente organizacional, seja na avaliação efetiva dos níveis de serviço da TI e na identificação de necessidades e problemas recorrentes, para implementação de ações corretivas capazes de produzir ganhos contínuos de qualidade dos serviços necessários à instituição pesquisada.

7 CONSIDERAÇÕES FINAIS

Neste trabalho evidenciou-se a aplicabilidade do processo KDD como ferramenta gerencial de TI hospitalar. Os resultados gerados pela sua aplicação, na base de dados de ordens de serviços informática de uma instituição hospitalar, demonstraram como o processo produziu conhecimento útil para avaliação dos níveis de serviços de TI e também os principais problemas atendidos pelo setor, a partir de informações existentes em grandes volumes de dados.

Tal conhecimento pode auxiliar tanto o gestor de TI como o do hospital na compreensão da qualidade de serviços disponibilizados pelo setor a toda a instituição, indicando a necessidade de implementação, ou ajustes, das ações de melhores práticas na TI. Da mesma forma, permite identificar quais são os problemas mais relevantes vivenciados e direcionar melhor os esforços e investimentos de informática na instituição hospitalar. Sendo o setor de informática essencial ao suporte à operação de toda a instituição, certamente todas as ações de melhoria devem refletir em todos os demais setores, gerando maior eficiência

REFERÊNCIAS

Agrawal, R., & Srikant, R. (1994). *Fast algorithms for mining association rules*. Acesso em 25 de junho de 2013, disponível em <http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf>

Arno, S. (2004). *Técnicas de Mineração de Dados*. Acesso em 11 de junho de 2013, disponível em <http://www.lsi.ufu.br/documentos/publicacoes/ano/2004/JAI-cap5.pdf>

Barbosa, J. M., Carneiro, T. de, & Tavares, A. I. [s.d.]. *Métodos de Classificação por Árvores de Decisão*. Acesso em 11 de setembro de 2013, disponível em UFOP – Universidade Federal de Ouro Preto: <http://www.decom.ufop.br/menotti/paa111/files/PCC104-111-ars-11.1-JulianaMoreiraBarbosa.pdf>

Becker, J., & Knackstedt, R. (2009). *Developing Maturity Models for IT Management – A Procedure*

operacional, confiabilidade e segurança aos serviços administrativos e também assistenciais.

Cabe destacar que o processo KDD não é único e pode ser trabalhado como elemento complementar a diversas outras metodologias, técnicas e ferramentas de gestão de serviços de TI.

Ainda assim, existe a necessidade de aprofundamento deste trabalho, no sentido de tratar diversos fatores como a redução de imperfeições das bases de dados originais, a inclusão de novos atributos que possam contribuir na construção das classes, homogeneização dos conteúdos de alguns atributos importantes como CAUSA_DANO e TIPO_SOLUCAO, entre outros. O ganho de qualidade do conteúdo e a melhor estruturação da base de dados pode permitir a produção de conhecimento cada vez mais efetivo e, por consequência, ampliar a possibilidade de atuação junto ao serviço de TI da instituição.

Com a ampliação do uso de sistemas ERP pelas instituições hospitalares, cada vez mais e maiores quantidades de informação estão sendo inseridas nos seus repositórios de dados, tornando cada vez mais importante a utilização de ferramentas capazes de manipular eficientemente grandes volumes de dados. Ser capaz de usufruir de todo o conhecimento armazenado dá às organizações vantagem competitiva num mercado cada vez mais acirrado.

O KDD e a mineração de dados, que vêm demonstrando sua aplicabilidade em diversos segmentos, demonstram ser úteis também como suporte à gestão de TI dos hospitais. A extração de conhecimento de bases bem estruturadas dá melhores condições de tomada de decisão aos seus gestores e maior qualidade de governança do setor.

Model and its Application. *Business & Information Systems Engineering*, 3.

Bordoloi, C. (2012). *Is Business Performance Improved by Better IT?* Acesso em 15 de maio de 2013, disponível em Agile IT governance: <http://agileitgovernance.com/2012/07/06/is-business-performance-improved-by-better-it-governance>

Burmester, H. (2012). *Manual de Gestão Hospitalar* (1. ed.). Rio de Janeiro, RJ, Brasil: Editora FGV.

Campanairio, M. A., Maccari, E. A., & Paulo, W. C. (2004). ERP – Enterprise Resource Planning “Como escolher o melhor para uma empresa”. *RAI – Revista de Administração e Inovação*, 1 (2), pp. 5-19.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in Databases*. Acesso em 13 de junho de 2014, disponível em kdnuggets.com

<http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>

Frank, E., & Witten, I. H. (1998). *Generating accurate rule sets without global optimization*. Acesso em 1º de junho de 2013, disponível em citeseerx: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.47.6014>

Galvão, N. D., & Marin, H. de F. (2009). *Técnica de mineração de dados: uma revisão da literatura*. Acesso em 1º de junho de 2013, disponível em scielo: <http://www.scielo.br/pdf/ape/v22n5/14.pdf>

Gnanapriya, S. S., Sumitha, R., & Kumar, S. M. (2010). Data Mining Concepts and Techniques. *CiiT International Journal of Data Mining and Knowledge Engineering*, 2 (9).

Goldschmidt, R., & Passos, E. (2005). *Data Mining um guia prático*. Rio de Janeiro, RJ, Brasil: Campus.

Gutierrez, M. A. (2011). Sistemas de Informação Hospitalares: progressos e avanços. *Journal of Health Informatics*, 3 (2).

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2. ed.). São Francisco, CA, EUA: Morgan Kaufmann.

Hannah, K. J., Ball, J. M., & Edwards, M. J. (2009). *Introdução à informática em enfermagem*. Porto Alegre, RS, Brasil: Artmed.

Hauser, S. D. (2006). *Considerações sobre o trabalho psicopedagógico em ambiente hospitalar*. Acesso em 15 de junho de 2013, disponível em abpp: <http://www.abpp.com.br/artigos>

IT Governance Institute – ITGI. (2007). *COBIT 4.1*. Rolling Meadows, EUA: T Governance Institute – ITGI

IT Governance Institute. [s.d.]. *About IT Governance*. Acesso em 10 de outubro de 2013, disponível em <http://www.itgi.org/About-Governance-of-Enterprise-IT.html>

IT Governance Institute. (2007). *COBIT 4.1*. Rolling Meadows, EUA.

Larose, D. T. (2005). *Discovering knowledge in data. An introduction to data mining*. New Jersey, EUA: John Wiley & Sons, Inc.

Lee, J. J., & Ben-Natan, R. (2001). *Integrating Service Level Agreements Optimizing Your OSS for SLA Delivery*. Indianapolis, EUA: Wiley Publishing Inc.

Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (2. ed.). New York, USA: Springer.

Malagón-Londoño, G., Morera, R. G., & Laverde, G. P. (2010). *Administração hospitalar* (3. ed.). (A. D. Paulo, Trad.) Buenos Aires, Argentina: Editorial Médica Panamericana S.A.C.F.

Malferrari, C. J., Santos, D. S., Lodi, J. B., Saraiva, J. W., Vieira, L. C., Tuccori, P. et al. (1983). *O Hospital e a visão administrativa contemporânea*. São Paulo, SP, Brasil: Livraria Pioneira Editora.

Olson, D. L., & Delen, D. (2008). *Advanced Data Mining Techniques*. Springer, EUA: Berlin Heidelberg.

Pereira, S. R., Paiva, P. B., Souza, P. S., Siqueira, G., & Pereira, A. R. (1º de novembro de 2012). Sistemas de Informação para Gestão Hospitalar. *Journal of Health Informatics*, 4 (4), pp. 170-5.

Pinochet, L. H. (2011). Tendências de Tecnologia de Informação na Gestão da Saúde. *Mundo da Saúde*, 34 (4), 382-394.

Rezende, S. O. (2003). *Sistemas inteligentes: fundamentos e aplicações*. Barueri, São Paulo, Brasil: Manole.

Robinson, N. (2007). *The Many Faces of IT Governance: Crafting an IT Governance Architecture*. Acesso em 1º de junho de 2013, disponível em <http://www.isaca.org/Journal/Past-Issues/2007/Volume-1/Documents/jpdf-0701-The-Many-Faces.pdf>

Salu, E. J. (2013). *Administração hospitalar no Brasil*. Barueri, São Paulo, Brasil: Manole.

Senhoras, E. M. (Jan.-jun. de 2007). A cultura na organização hospitalar e as políticas culturais de coordenação de comunicação e aprendizagem. *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, 01 (01), pp. 45-55.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3. ed.). Burlington, MA, USA: Morgan Kaufmann Publishers.