

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/241704588>

# Projeto de módulo de Data Mining para Scout Voleibol

Article · November 2008

CITATIONS

0

READS

367

5 authors, including:



**Emerson Butzen**

Universidade do Vale do Rio dos Sinos

7 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)



**Alexandre de Oliveira Zamberlam**

7 PUBLICATIONS 4 CITATIONS

[SEE PROFILE](#)



**Ricardo Ferreira de Oliveira**

Universidade Feevale

3 PUBLICATIONS 0 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



IA na quadra de vôlei [View project](#)



Polymorphisms in cytochrome P450 2E1 (CYP2E1): effects on the three-dimensional structure, molecular dynamics and function [View project](#)

All content following this page was uploaded by **Emerson Butzen** on 31 May 2014.

The user has requested enhancement of the downloaded file.

# Projeto de módulo de *Data Mining* para Scout Voleibol

Émerson B. Marques<sup>1</sup>, Alexandre de O. Zamberlam<sup>1</sup>, Ricardo F. de Oliveira<sup>1</sup>,  
Luis H. Raimann<sup>1</sup> e Luciano Vaghetti de Oliveira<sup>2</sup>

<sup>1</sup>Centro Universitário Feevale – Instituto de Ciências Exatas e Tecnológicas - Novo  
Hamburgo – RS – Brasil

<sup>2</sup>Universidade Luterana do Brasil – Curso de Educação Física - Canoas – RS – Brasil

{butzen, alexz, ricardofo}@feevale.br,  
{luisraimann, vaghetti}@gmail.com

**Abstract.** *This paper describes the research made about statistical monitoring systems (scout) aided by the technology of data mining. These systems are used to evaluate athletes performance in volleyball courts, allowing to the staff technical to decide on training, strategies of game and tactics on plays.*

**Resumo.** *Este artigo descreve a pesquisa realizada, até agora, sobre sistemas de monitoramento estatístico de desempenho de atletas de voleibol (scout). O presente texto aborda o projeto de um módulo de data mining que integre o sistema scout, possibilitando às comissões técnicas realizarem ações mais concretas sobre estratégias e táticas de jogo.*

## 1. Introdução

Sistemas de Apoio a Decisão, normalmente utilizados em empresas no processo de tomada de decisão, também são utilizados em equipes de voleibol e/ou em outros esportes, com a finalidade de gerar melhores resultados. Nesse universo de equipes esportivas, foram elaborados sistemas estatísticos de monitoramento de desempenho de atletas e de equipes, também conhecidos como sistemas de *scout*, permitindo às comissões técnicas decidirem sobre treinamentos, estratégias de jogo e táticas sobre jogadas. Contudo, a quantia de informações (precisas e/ou desnecessárias) é assim mesmo enorme, deixando o processo de decidir bastante empírico, ou seja, dependente da experiência do técnico ou de algum integrante da comissão técnica.

A mineração de dados (*data mining*) entra nesse contexto por oferecer um ambiente propício e dados mais confiáveis, pois através dela, há processos de seleção, de exploração e de modelagem de dados em grande escala (GIUDICI, 2003).

Por meio da base de dados do sistema de *scout*, construído por Raimann (2008), está se projetando um módulo de *data mining* que aplica, sucessivamente, as técnicas de mineração de dados previamente escolhidas: regras de associação, agrupamento, dados em séries temporais e padrões sequenciais.

Dessa forma, este texto pretende apresentar as atividades já realizadas. Assim, dividiu-se o texto em 5 seções. A seção 2 apresenta conceitos e técnicas de mineração de dados usados no trabalho. A seção 3 descreve, brevemente, um sistema *scout*. Na seção 4, é descrito o módulo de mineração. Finalmente, algumas considerações sobre este trabalho são apresentadas.

## 2. Mineração de Dados

A mineração de dados é uma etapa do processo de Descoberta do Conhecimento em Banco de Dados mais conhecido por *Knowledge Discovery in Databases* (KDD). Muitos autores escrevem que muitas vezes o termo mineração de dados é utilizado como sinônimo para todo o processo de descoberta de conhecimento, entretanto ela é a etapa mais importante, respondendo por 15 a 25 por cento do processo de descoberta. Enfim, para muitos autores, a mineração de dados é o processo de extrair informações válidas antes desconhecidas, de grandes bases de dados, auxiliando em decisões cruciais no mundo dos negócios. Para isso, *data mining* utiliza-se de técnicas ou algoritmos de áreas como Aprendizado de Máquinas, Estatística, Redes Neurais, Algoritmos Genéricos, etc (ELMASRI; NAVATHE, 2005). O mesmo autor aponta que a mineração de dados apóia o conhecimento indutivo, descobrindo novas regras e padrões nos dados minerados.

Como a mineração de dados é uma parte do processo de Descoberta do Conhecimento em Banco de Dados, há etapas que devem ser realizadas antes de aplicar-se qualquer técnica de *data mining*. Conforme Bramer (2007), os dados possivelmente vêm a partir de muitas fontes, sendo necessário que esses dados sejam integrados e armazenados em lugar comum. O próximo passo é a preparação dos dados, então é aplicada alguma técnica de mineração de dados, que produz uma saída na forma de regras ou em algum tipo de padrão. Essas regras ou padrões são também conhecidos como modos de descoberta.

Para Kantardzic (2003) o processo de mineração de dados é composto por cinco fases: definição do problema; seleção, coleta dos dados; pré-processamento dos dados; especificação de possível método; interpretação e análise dos dados produzidos pelo método. Já os métodos ou técnicas de mineração de dados são na verdade algoritmos computacionais. Cada um desses algoritmos tem características particulares, normalmente entradas e saídas específicas. Neste trabalho vem se utilizando regras de associação, classificação, agrupamento, padrões sequenciais e padrões em séries temporais.

### 2.1. Técnica Regra de Associação (*Association Rule*)

Através dela é possível recuperar todos os padrões interessantes em uma base de dados. A base de dados é uma coleção de transações. No voleibol há características semelhantes que podem ser identificadas através de regras de associação. A Tabela 1 mostra os jogadores em quadra e o aproveitamento da equipe no fundamento bloqueio.

**Tabela 1 – Jogos de jogadores e aproveitamento do fundamento bloqueio**

IdJogo (Partida)	iNumeroCamisetaJogador (em quadra)	Bloqueio (Resultado)
1	2, 5, 6, 8, 10, 11	Excelente
1	1, 4, 5, 6, 8, 10	Regular
2	2, 3, 5, 8, 10, 11	Bom
2	1, 3, 4, 8, 9, 11	Péssimo

Uma regra de associação é da forma  $X \Rightarrow Y$ , onde  $X = \{x_1, x_2, \dots, x_n\}$  e  $Y = \{y_1, y_2, \dots, y_m\}$  são conjuntos de itens com  $x_i$  e  $y_j$  sendo distintos para todo  $i$  e todo  $j$ . Essa associação estabelece que quando o jogador  $X$  estiver escalado, ele estará propenso a jogar com o Jogador  $Y$ , caso também estiver (na mesma equipe ou da equipe adversária, um determinado bloqueador escalado para bloquear específico atacante). Normalmente

essas regras de associação têm a forma LME (lado da mão esquerda) e LMD (lado da mão direita), cada uma formando um conjunto de itens. Quando LME une-se a LMD ( $LME \cup LMD$ ) tem-se o conjunto-item, que é, no exemplo do voleibol, o conjunto de todos os jogadores escalados. Para descobrir interesses nessa regra, ela precisa satisfazer algumas medidas, as mais comuns fornecem suporte e confiança (ELMASRI; NAVATHE, 2005). O suporte para a regra  $LME \Rightarrow LMD$  refere-se à frequência que ela ocorre no banco de dados, ou seja, o percentual de transações que contém todos os itens na própria relação  $LME \cup LMD$ . Se o suporte é baixo, implica que não existe evidência significativa que os itens  $LME \cup LMD$  ocorram juntos. A confiança dessa regra é calculada como o suporte ( $LME \cup LMD$ ) / suporte (LMD). Pode-se assim analisar a possibilidade dos jogadores do LMD sejam escalados juntos, dado que os jogadores do LME também sejam escalados pelo técnico (ELMASRI; NAVATHE, 2005). Para exemplificar o suporte e a confiança, considere essas regras: Jogador 10  $\Rightarrow$  Jogador 11 e Jogador 8  $\Rightarrow$  Jogador 11. Observa-se na Tabela 1 que existem quatro transações de duas partidas de vôlei: o suporte para {Jogador 10, Jogador 11} é de 50%, e o suporte de {Jogador 8, Jogador 11} é 75%. Já a confiança de Jogador 10  $\Rightarrow$  Jogador 11 é de 66,7%, ou seja, das três vezes em que o jogador 10 é escalado, duas tem a companhia do jogador 11, e a confiança do Jogador 8  $\Rightarrow$  Jogador 11 é também de 66,7%.

Pode-se notar que o suporte e a confiança nem sempre são proporcionais, a meta para as regras de associação é gerar todas as possíveis regras que excedam a especificação do usuário garantindo suporte e confiança acima do limite definido (ELMASRI; NAVATHE, 2005). O principal algoritmo de regras de associação é o algoritmo *Apriori*.

## 2.2. Técnica Agrupamento (*Clustering*)

Uma dada população de eventos ou novos itens podem ser particionados (segmentados) em conjuntos de elementos ‘padrões’ (ELMASRI; NAVATHE, 2005). Bramer (2007) confirma que os algoritmos de agrupamento analisam os dados para encontrar grupos de itens que são semelhantes. Amostras de agrupamento são representadas como um vetor de medições, ou, mais formalmente, como um ponto em um espaço multidimensional. As amostras de um agrupamento válido são mais semelhantes (não necessariamente iguais) entre si do que as amostras que pertencem a um agrupamento diferente (KANTARDZIC, 2003). A Tabela 2 tem um simples exemplo de agrupamento de informação. São doze atletas distribuídos em três agrupamentos. As características que descrevem esses atletas são: o aproveitamento em percentual no fundamento bloqueio e o número de bloqueios realizados.

Pode-se concluir que o Agrupamento 1 tem o melhor aproveitamento, mas pouquíssimos bloqueios foram efetuados por seus atletas. No Agrupamento 2, existe a melhor média, pois vários bloqueios foram efetuados. O Agrupamento 3 apresenta atletas que possivelmente não tenham esse fundamento como principal característica.

Para utilizar a técnica de agrupamento, é necessário um critério objetivo. A entrada para uma análise de agrupamento pode ser descrita como um par requisitado ( $X, s$ ), ou ( $X, d$ ), onde  $X$  é um jogo (descrições das amostras), e  $s$  e  $d$  são medidas para a similaridade ou a dissimilaridade (distância) entre amostras, respectivamente. A saída é uma divisão  $\Lambda = \{G_1, G_2, \dots, G_N\}$  onde  $G_k$ ,  $k = 1, \dots, N$  é um subconjunto cristalino de  $X$  tal que  $G_1 \cup G_2 \cup \dots \cup G_N = X$ , e  $G_i \cap G_j = \emptyset$ ,  $i \neq j$ . Assim os membros  $G_1, G_2, \dots, G_N$  de  $\Lambda$  são chamados de agrupamentos. Cada um dos agrupamentos é descrito com

algumas características. Na descoberta, a base de agrupamentos, tanto o agrupamento (um conjunto de pontos distintos em X), quanto suas características, são gerados como resultado de um procedimento de agrupamento (KANTARDZIC, 2003).

**Tabela 2 – Simplex exemplo de agrupamento de informação**

Agrupamentos	Aproveitamento	Bloqueios Realizados	Atletas
Agrupamento 1	100%	1	Jogador 1
	90%	4	Jogador 3
	80%	1	Jogador 6
Agrupamento 2	70%	20	Jogador 2, Jogador 9
	60%	30	Jogador 4, Jogador 5
	50%	30	Jogador 8
Agrupamento 3	40%	10	Jogador 7
	20%	3	Jogador 11
	0%	5	Jogador 10, Jogador 12

A vasta coleção de algoritmos de agrupamento disponíveis na literatura e os diferentes softwares confundem os usuários que tentam selecionar aproximações mais apropriadas para o seu problema. A maior parte dos algoritmos de agrupamento tem por base as duas seguintes abordagens: agrupamento hierárquico e agrupamento particional iterativo.

### 2.3. Técnica Dados em Séries Temporais (*Time-Series Data*)

Para Elmasri e Navathe (2005), esses padrões podem ser encontrados em posições de uma série temporal de dados, que é uma sequência de dados capturada a intervalos regulares (segundos, horas, dias, semanas, etc). Han e Kamber (2006) confirmam, escrevendo que séries temporais consistem em sequências de valores ou medidas repetidas, excessivamente, em intervalos de tempos.

No voleibol, pode-se analisar o desempenho de atletas pelo tempo de jogo, e analisar assim qual é o tempo de jogo ideal para ele (período no *set* ou no jogo todo). Existem atletas que jogam com um mesmo nível de desempenho a partida inteira, outros atletas têm um início de alto desempenho, e no decorrer da partida o nível cai. A de se considerar, aqui, também o aspecto emocional de cada atleta, não apenas os números. Na Tabela 3, é possível observar atletas e seus níveis de desempenho nas fases de um *set*. Esse é um exemplo ilustrativo apenas, já que em uma situação real, o volume de dados seria muito maior. Pode-se, também, dividir um *set* em fases (de 0 a 8 pontos, fase um; de 9 a 19, fase dois; e de 20 a 25, fase 3). Nessa divisão é possível perceber os diferentes comportamentos dos atletas (qualidade da jogada, concentração, obediência técnica e tática) devido ao aumento do estresse emocional do final do *set*.

**Tabela 3 – Aproveitamento de Saque (Atleta/Tempo de Jogo)**

Atletas	Tempo de Jogo (fases)		
	Fase 1	Fase 2	Fase 3
Jogador 01	100 %	80 %	40 %
Jogador 02	60 %	60 %	70 %

Em geral, são dois os objetivos na análise de séries temporais: modelagem das séries temporais (isto é, ganhar uma visão sobre os mecanismos subjacentes ou forças que geram as séries temporais), e previsão de séries temporais (isto é, prever os futuros valores do tempo de séries de variáveis), as técnicas mais usadas são a análise de tendências e a pesquisa por semelhança (HAN; KAMBER, 2006).

No voleibol, a análise de tendência parece ser a mais atrativa, já que a partir dos dados históricos é possível traçar tendências de jogadas de equipes adversárias, e analisando os dados da Tabela 3, pode-se visualizar a tendência de aproveitamento dos atletas conforme as fases de jogo.

## 2.4. Padrões seqüenciais (*Sequential Patterns*)

Elmasri e Navathe (2005) apontam que a técnica de padrões seqüenciais é a investigação de seqüências de ações ou eventos. Han e Kamber (2006) escrevem que é uma técnica desafiadora, pois pode gerar e/ou testar um número combinatório explosivo de seqüências intermediárias. Essa técnica tem algumas similaridades com a técnica de regras, a diferença é que faz exame da dimensão seqüencial dos dados analisados.

No voleibol, essa técnica pode ser utilizada para identificar o comportamento padrão de ataque ou defesa de equipes adversárias. Existem casos em que equipes têm jogadas prontas (ensaiadas), principalmente por características e/ou limitações técnicas de seus atletas. Essas jogadas podem dar resultados positivos sempre que bem efetuadas, mas se a outra equipe identificar esse padrão de jogo durante a partida, pode se beneficiar, pois estaria evitando eficazmente a jogada de seu adversário. Mas muitas vezes não é identificada essa seqüência padrão de acontecimentos. Utilizando a técnica de padrões seqüências, esse comportamento padrão seria percebido no resultado da mineração.

## 3. SCOUT Voleibol

Conforme Bernardinho (2006), *scout* serve para capturar e processar informações estatísticas de desempenho dos atletas da equipe e da equipe adversária, monitorando assim desempenhos de atletas nos fundamentos (saque, defesa, ataque, bloqueio, recepção). Com isso, é possível auxiliar na tomada de decisão, na elaboração de planos técnicos (saque, levantamento, cortada, etc) e táticos (como evitar os tipos de jogadas do adversário). Em caráter ilustrativo, a Figura 1 apresenta a estatística de um jogo monitorado pelo *scout* de Roberta Giglio na Seleção Brasileira de Voleibol.

Competição: Jogos Olímpicos		Local: Grácia	
Nº do jogo: 8		Horário: 14:30h	
Jogo: Brasil X Itália		Placar: 3 x 1	
Data: 29/08/2004			

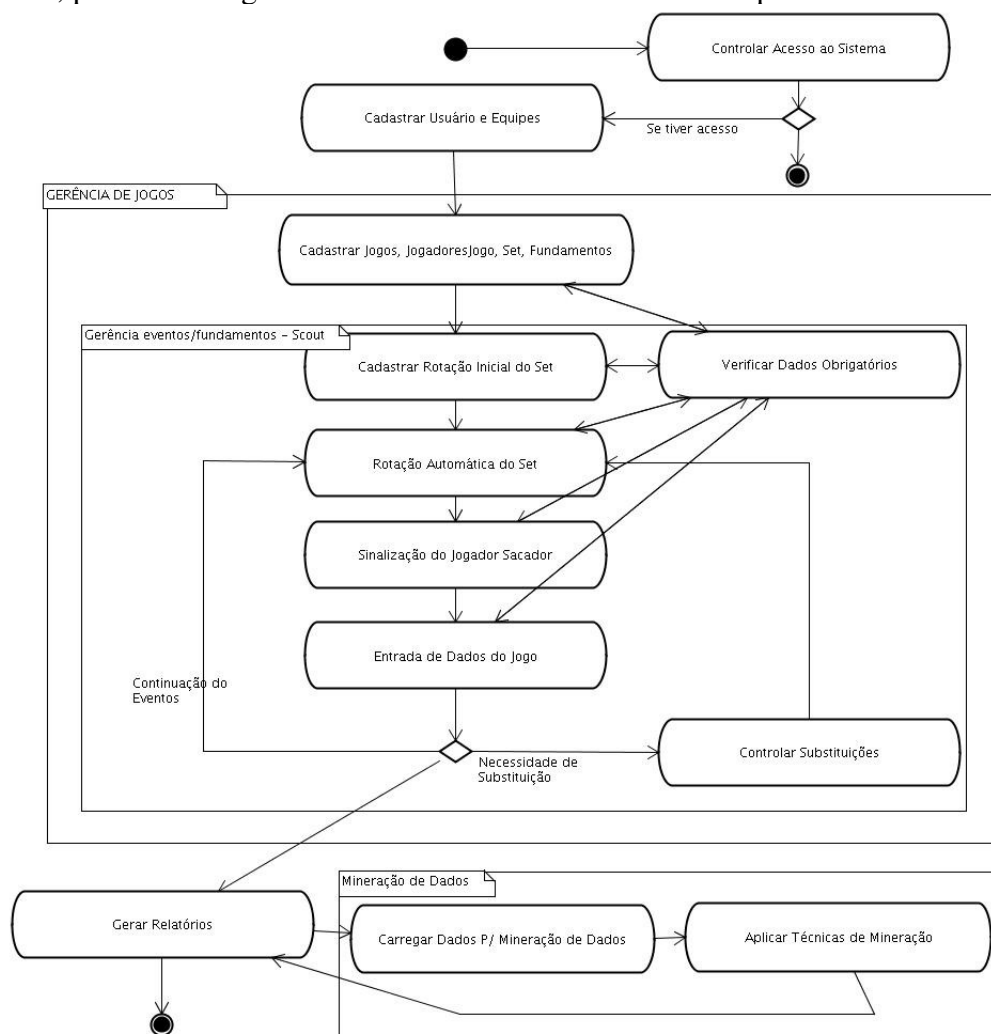
FUNDAMENTOS	SAQUE						BLOK		PASSE						CONTRA						ATAQUE						C + A		DEF	E	APV	
Atletas	SP	SD	SE	SF	Tot	%	BP	BC	PA	PC	PD	PP	Tot	%	CP	C	CN	CB	Tot	%	AC	AD	AF	AB	AE	Tot	%	Tot	%	DP	DN	Pto
3. Giovane															2	1			3	67	7		3			10	70	13	69			9
4. André H.			3	1	13	17	18	1	3																							
6. Mauricio																																
7. Giba	3	4	4	12	23	30	2	1	12	1	3	2	18	67	4	5	2		11	36	12	1	2	4	2	21	57	32	50	5	1	6
9. Canha	1		3	3	7	14	1	3							4	4			8	50	5	1	4	3	1	14	36	22	41			4
10. Escadinha									19		3	1	23	83																2	1	-1
11. Anderson			2		1	3	67	1									1		1	0	2			2		4	50	5	40			
12. Nalbert					1	1	0																									
13. Gustavo	3	5	1	5	14	57	2	3							1				1	100	9					9	100	10	100			13
14. Rodrigo																																
17. Ricardo			9		10	19	47	1															1			1	0	1				1
18. Dante	2	3	1	7	13	38	3	2	22	2	5	1	30	73	3	2			5	60	5		1	3	1	10	50	15	53	3	1	6
Equipe	9	26	10	52	97	36	11	12	53	3	11	4	71	75	14	12	3		29	48	40	2	11	12	4	69	58	98	55	11	4	48

Figura 1 – Planilha de aproveitamento (BERNARDINHO, 2006).

Entretanto, há grande quantidade de dados gerados, muitas vezes desnecessários, dificultando a visualização de padrões de jogo. Isso também força que a análise seja realizada após as partidas.

#### 4. O módulo proposto

O módulo de mineração interage com a base de dados, Figura 2, fornecendo informações mais precisas à comissão técnica, evitando assim o uso de complicadas planilhas, pois existem grandes dificuldades nas leituras dessas planilhas.



**Figura 2. Diagrama de Atividades Sistema Scout e módulo de mineração.**

Na Figura 3 é possível visualizar parte da estrutura do banco de dados do sistema de *scout* desenvolvido por Raimann (2008), a fim de exibir em quais dados a mineração deverá atuar. Através dessa estrutura do banco de dados, é possível perceber a quantidade de dados que uma partida de voleibol pode gerar. Após a partida ou *set* é que o módulo de mineração de dados entrará em ação. O módulo tem como recurso gerar relatórios para a comissão técnica. Por exemplo, relatórios: *i*) de jogadas de uma equipe de vôlei em um *set*, com base na análise de jogadas de um jogo completo, com o uso da técnica de regras de associação; *ii*) com agrupamento de dados dos fundamentos de saque e bloqueio, coletados em jogos de uma equipe e organizados em faixas de aproveitamento - técnica de agrupamento; *iii*) de um comportamento padrão (seqüência

de jogadas) da defesa de uma equipe de vôlei adversária após um saque da outra equipe - técnica de padrões seqüências; iv) de padrões de comportamentos de equipes de voleibol após determinada faixa de pontuação (aspecto emocional dos atletas), técnica de padrões com séries temporais.

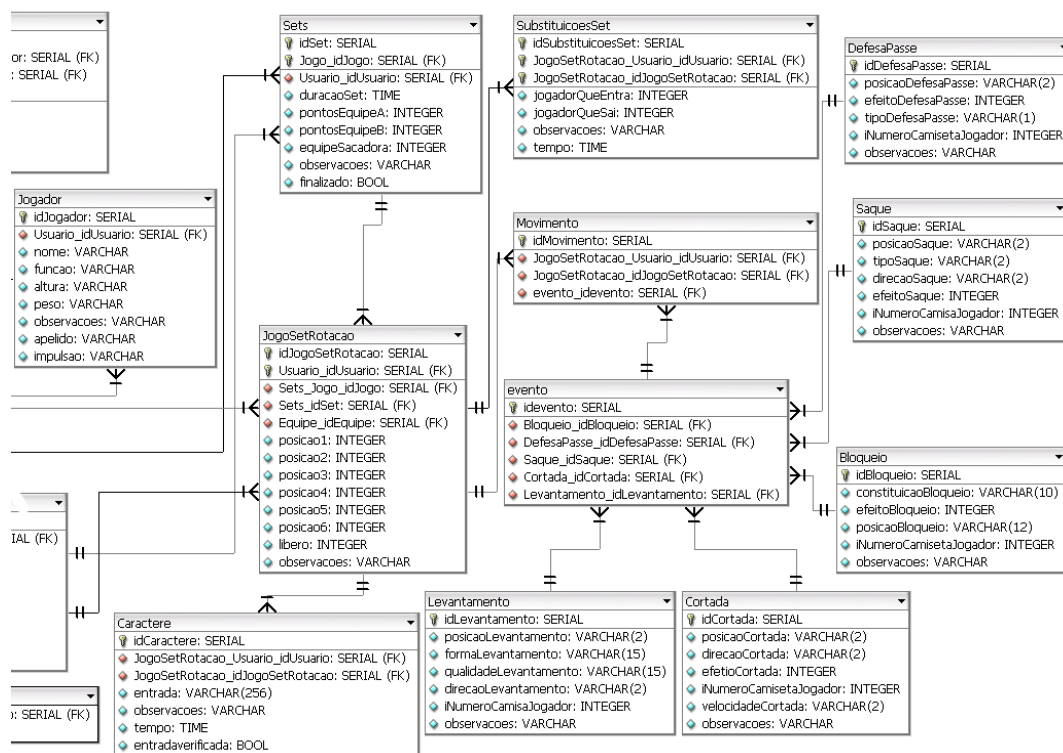
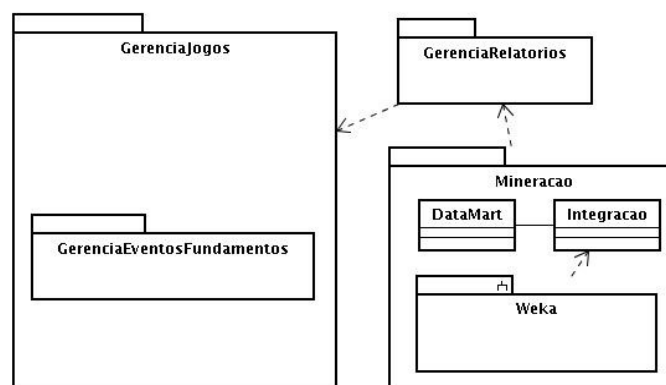


Figura 3. ER do sistema *scout* (RAIMANN, 2008)

Essas situações mostram informações importantes para uma comissão técnica de uma equipe de vôlei, pois são resultados que podem ser percebidos durante a partida e não após, como acontece geralmente nos sistemas de *scout*. Em alguns sistemas, essas informações são levantadas manualmente, após isso, apresentadas às suas comissões técnicas.

O módulo de mineração interage com a ferramenta WEKA (*Waikato Enviroment Knowledge Analysis*), que é uma coleção de algoritmos da aprendizagem de máquina para tarefas de mineração de dados. Esses algoritmos podem ser aplicados diretamente a uma série de dados ou serem chamados de seu próprio código Java. O Weka contém ferramentas para o pré-processamento dos dados, a classificação, a regressão, o agrupamento, as regras da associação e visualização, é também bem adequado para o desenvolvimento de novos sistemas de aprendizagem (WITTEN; FRANK, 2005). Para trabalhar as séries de dados, o módulo de mineração proposto converte essa série em arquivos do formato ARFF (*Attribute-Relation File Format*). Esse formato é um arquivo texto composto por três partes: relação, atributos e dados. Outra característica do WEKA é sua portabilidade para outras aplicações Java através de suas classes. Assim sendo, como o sistema proposto por Raimann (2008) foi projetado em Java, a integração com WEKA é facilitada. Observe a Figura 4 que apresenta a idéia dessa integração.





**Figura 4. Diagrama de Pacotes da integração.**

## 5. Considerações finais

Com esse sistema e esse módulo de *data mining*, espera-se que equipes tenham melhores condições de desenvolver os fundamentos ineficientes de cada atleta, e obter estratégias de jogo num menor tempo, auxiliando a comissão técnica na sua tomada de decisão, elevando assim a qualidade do voleibol praticado pelas equipes.

O módulo proposto está em fase de finalização da implementação, no que se refere a integração com a base e a ferramenta WEKA. Em breve, iniciarão os primeiros testes para verificar a eficácia do módulo junto ao *scout* desenvolvido por Raimann (2008).

## Referências bibliográficas

- BERNARDINHO. (2006) Transformando suor em ouro. Rio de Janeiro, RJ: Sextante.
- BRAMER, Max. (2007) Principles of Data Mining (Undergraduate Topics in Computer Science). Londres, Inglaterra: Springer-Verlag London Ltd.
- ELMASRI, Ramez; NAVATHE, Shamkant B. (2005) Sistemas de banco de dados. 4. ed. São Paulo, SP: Pearson Addison Wesley.
- GIUDICI, Paolo. (2003) Applied Data Mining: Statistical Methods for Business and Industry. Chichester, West Sussex, Inglaterra: John Wiley & Sons Ltd.
- HAN, Jiawei; KAMBER, Micheline, (2006) Data Mining Concepts and Techniques. São Francisco, Califórnia, EUA: Morgan Kaufman Publishers.
- KANTARDZIC, Mehmed. (2003) Data Mining: Concepts, Models, Methods, and Algorithms. Hoboken, Nova Jersey, EUA: John Wiley & Sons Inc.
- RAIMANN, Luís Henrique, (2008) Scout: Sistema de monitoração em equipes de voleibol. Novo Hamburgo, RS: 2007. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Instituto de Ciências Exatas e Tecnológicas, Feevale. Disponível em: < <http://nead.feevale.br/tc/files/1247.pdf>>. Acesso em: 10 mar 2008.
- ZAMBERLAM, Alexandre de Oliveira; WIVES, Leandro Krug; GOULART, Rodrigo Rafael Villarreal; SILVEIRA, Roni Gilberto. (2005) A IA entrando na quadra de vôlei: scout inteligente. Hifen, Uruguaiana, RS , v.29, n.55/56.
- WITTEN, Ian H.; FRANK, Eibe. Data Mining: Practical machine learning tools and techniques. 2<sup>a</sup> ed. São Francisco, Califórnia, EUA: Morgan Kaufman Publishers, 2005. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/index.html>>. Acesso em: 24 jun. 2008.