

ANÁLISE DO ALGORITMO NAIVE BAYES NA CLASSIFICAÇÃO DE AMOSTRAS DO BANCO DE DADOS HEPATITE

Larissa Duarte de Britto Lira ¹
Ingrid Rafaella dos Santos Melo ²
Herbert Schafer Neto ³
Kátia Suely Queiroz Silva Ribeiro ⁴

RESUMO

Um médico conclui o diagnóstico de uma doença avaliando a condição do paciente ou comparando com outro paciente que tenha condições ou sintomas semelhantes. Com base na base de dados da UCI existem 17 sintomas de hepatite que podem ser semelhantes a outras doenças, por isso precisa de um método para encontrar os principais sintomas. Na ciência da computação, essa tarefa pode ser feita por um programa de computador que inclua um algoritmo inteligente. Técnicas de mineração de dados são amplamente utilizadas na classificação e previsão no campo da bioinformática para análise de dados biomédicos. O objetivo do estudo é Avaliar o desempenho do modelo de suporte a decisão baseado no algoritmo Naive Bayes para análise de dados prognósticos da hepatite. O estudo conclui que o desempenho da classificação de Naive Bayes é eficaz para a classificação do conjunto de dados de hepatite.

Palavras-chave: Hepatite, Mineração de dados, Naive Bayes.

INTRODUÇÃO

A hepatite é um processo inflamatório que acomete o fígado podendo ser causado por diversas etiologias, infecciosas ou não, cujas manifestações clínicas podem variar de acordo com o agente causal, porém a icterícia é o sinal mais frequente e o que mais chama a atenção pela tonalidade amarelada da pele e olhos (Jayakumar et al. 2013; Poddar et al. 2013). Cinco espécies virais são responsáveis pela maior parte dos episódios de hepatite aguda e de hepatite crônica: vírus da hepatite A (HAV), vírus da hepatite B (HBV), vírus da hepatite C (HCV), vírus da hepatite D (HDV) – viroide defeituoso revestido pelo antígeno da hepatite B, e vírus da hepatite E (HEV).

¹ Mestranda em Modelos de Decisão e Saúde da Universidade Federal da Paraíba- UFPB, larissadblira@hotmail.com;

² Mestranda em Modelos de Decisão e Saúde da Universidade Federal da Paraíba- UFPB, ingridmeello@gmail.com;

³ Mestrando em Ciência da Educação da Universidade Lusófona de Humanidades e Tecnologia- ULHT, neto_2@hotmail.com.

⁴ Professor da Pós-graduação em Modelos de Decisão e Saúde da Universidade Federal da Paraíba- UFPB, katiaribeiro.ufpb@gmail.com.

Técnicas de mineração de dados são amplamente utilizadas para a classificação de dados. A classificação consiste no processo de encontrar, através de aprendizado de máquina, um modelo ou função que descreva diferentes classes de dados (Han; Kamber 2006). O objetivo da classificação é rotular, automaticamente, novas instâncias da base de dados com uma determinada classe aplicando o modelo ou função “aprendidos”. Este modelo é baseado no valor dos atributos das instâncias de treinamento. Diversos classificadores foram propostos nos últimos anos, alguns utilizam árvores de decisão para rotular registros e outros algoritmos se baseiam em redes neurais artificiais, modelos probabilísticos (bayesianos) ou em regras (Mitchell 1997).

Enquanto que uma pessoa física pode analisar um problema e escolher alternativas de decisão de maneira informal, na saúde este processo deve ser metodologicamente estruturado, com a finalidade de reduzir erros de diagnóstico e prescrições, além do aperfeiçoamento e ampliação do conhecimento das equipes através do levantamento de informações relacionadas ao atendimento às populações (Passaia; Pilla, 2010).

O estudo tem o intuito de avaliar o desempenho do modelo de suporte a decisão baseado no algoritmo Naive Bayes para o banco de Dados Hepatite, com intuito de prever a presença ou ausência do vírus da hepatite, além de fornecer uma descrição do prognóstico. Sendo realizado para justificar o desempenho de métodos conjuntos em um banco de dados médicos que podem ser usado para obtenção de diagnósticos eficazes, que por sua vez aumentariam o índice de saúde.

Atualmente, a tecnologia e a base de dados de computadores ajudam nessa tarefa pois coletam e armazenam uma enorme quantidade de dados. O grande tamanho da maioria das bases de dados impossibilita o ser humano de interpretar dados. Portanto, os computadores são necessários para extrair resultados, minimizando erros.

METODOLOGIA

3.1 Conjunto de dados de Hepatite

O conjunto de dados da doença hepatite foi obtido do UCI Machine Learning Repository. Os dados possuíam 155 registros com 19 atributos e um 1 atributo de classe. Dos 155 dados do registro, 32 registros rotularam óbito e 123 registros rotularam vivo. A descrição do atributo do conjunto de dados de hepatite é descrita na Tabela 1. Os dados incluem a

(83) 3322.3222

contato@conapesc.com.br

www.conapesc.com.br

descrição de prognóstico e preveem se um paciente está infectado com Hepatite de acordo com o resultado dos exames feitos no paciente.

TABELA 1 – INFORMAÇÕES DE DADOS DA DOENÇA HEPATITE

	ATRIBUTO	TIPO	VALORES
1.	Classe	Categórica	Óbito, vida
2.	Idade	Numérica	Valores Numéricos
3.	Sexo	Categórica	1 = Não, 2 = Sim
4.	Esteróide	Categórica	1 = Não, 2 = Sim
5.	Antiviral	Categórica	1 = Não, 2 = Sim
6.	Fadiga	Categórica	1 = Não, 2 = Sim
7.	Indisposição	Categórica	1 = Não, 2 = Sim
8.	Anorexia	Categórica	1 = Não, 2 = Sim
9.	Fígado crescido	Categórica	1 = Não, 2 = Sim
10.	Fígado firme	Categórica	1 = Não, 2 = Sim
11.	Baço palpável	Categórica	1 = Não, 2 = Sim
12.	Aranhas	Categórica	1 = Não, 2 = Sim
13.	Ascitas	Categórica	1 = Não, 2 = Sim
14.	Varizes	Categórica	1 = Não, 2 = Sim
15.	Bilirrubina	Numérica	0,39, 0,80, 1,20, 2,00, 3,00, 4,00
16.	Fosfato de alk	Numérica	33, 80, 120, 160, 200, 250
17.	TGO	Numérica	13, 100, 200, 300, 400, 500
18.	Albumina	Numérica	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
19.	Protíme	Numérica	10, 20, 30, 40, 50, 60, 70, 80, 90
20.	Histologia	Categórica	1 = Não, 2 = Sim

3.2 Mineração de Dados

A fase de mineração de dados é uma fase do processo de Descoberta de Conhecimento em Banco de Dados (DCBD). Esta etapa é responsável pela aplicação dos algoritmos que são capazes de identificar e extrair padrões relevantes presente nos dados (Han, 2001; Witten, 2000).

O processo DCBD é interdisciplinar, tanto em sua aplicação, quanto das suas fundamentações teóricas. O processo pode ser aplicado a qualquer problema de identificação de padrões em dados e contém fundamentação de diversas áreas como a banco de dados, inteligência artificial, estatística, probabilidade e visualização de dados.

O processo de diagnóstico começa selecionando os recursos do conjunto de dados usando as ferramentas de aplicação WEKA 3.8.2 (Waikato Environment for Knowledge Analysis) que é formado por um conjunto de implementações de algoritmos de diversas técnicas de Mineração de Dados (University of Waikato, 2010).

O Weka está implementado na linguagem Java, que tem como principal característica a sua portabilidade, desta forma é possível utiliza-la em diferentes sistemas operacionais, além de aproveitar os principais benefícios da orientação a objetos.

Para testes neste estudo será utilizado o algoritmo naive bayes, os dados serão testados por cross-validation e percentil Split. O cross-validation, uma técnica de avaliação padrão, é uma maneira sistemática de executar repetições percentuais repetidas. Divida um conjunto de dados em 10 partes (“dobras”), depois segure cada peça para testes e treine nos 9 restantes juntos. Isso dá 10 resultados de avaliação, que são calculados. No cross-validation “estratificado”, ao fazer a divisão inicial, asseguramos que cada dobra contenha aproximadamente a proporção correta dos valores da classe. Tendo feito uma validação cruzada de 10 vezes e computado os resultados da avaliação, Weka invoca o algoritmo de aprendizado em um tempo final (11) em todo o conjunto de dados para obter o modelo que ele imprime.

3.3 Redes Bayesianas

As Redes Bayesianas, também conhecidas como Redes casuais, Rede de crença e Gráficos de dependência probabilística, surgiram na década de 80 e têm sido aplicadas em uma grande variedade de atividades do mundo real (Bobbio et al., 2001). Algumas aplicações atuais se estendem às áreas como finanças, saúde, desenvolvimento de jogos (Vieira Filho e Albuquerque, 2007), entre outras.

A construção de uma Rede Bayesiana não é trivial, além de existir vários métodos para a estimação de estruturas de rede através do conjunto de dados, os métodos podem ser influenciados por fatores como a ordem e escolha das variáveis que compõem o problema. Esse problema proporciona atualmente intensas pesquisas buscando um método ótimo para estimação de estruturas DAG para domínios de problemas práticos.

As redes bayesianas podem representar relações causais entre variáveis contidas na estrutura de rede bayesiana. Por exemplo, uma rede bayesiana pode representar uma relação probabilística entre doença e sintomas, podem ser usadas para calcular a probabilidade da presença de vários sintomas da doença. As redes bayesianas podem fazer inferências probabilísticas. A inferência probabilística é prever o valor de variáveis desconhecidas diretamente usando os valores de outras variáveis conhecidas. A inferência probabilística pode ser feita se a Distribuição de Probabilidade de Junção de todas as variáveis conhecidas for conhecida (Krause, 1998). A inferência probabilística pode ser feita se a rede bayesiana tiver sido construída, então o que precisa ser feito primeiro é construir a estrutura da Rede Bayesiana (Santika, 2012).

3.4 Naive Bayes

É uma técnica de classificação baseada no teorema de Bayes com uma suposição de independência entre os preditores. Em termos simples, um classificador Naive Bayes assume que a presença de uma característica particular em uma classe não está relacionada com a presença de qualquer outro recurso. O modelo Naive Bayes é fácil de construir e particularmente útil para grandes volumes de dados. Além de simples, Naive Bayes é conhecido por ganhar de métodos de classificação altamente sofisticados.

Teorema de Bayes fornece uma forma de calcular a probabilidade posterior $P(C | X)$ a partir de $P(C)$, $P(x)$ e $P(X | c)$. Veja a equação abaixo:

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$

Diagram illustrating the components of the equation:

- $P(c | x)$ is labeled as "Probabilidade posterior" (Posterior Probability).
- $P(x | c)$ is labeled as "Probabilidade" (Probability).
- $P(c)$ is labeled as "Probabilidade original da Classe" (Original Class Probability).
- $P(x)$ is labeled as "Preditores da probabilidade posterior" (Predictors of posterior probability).

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Acima,

- $P(c | x)$ é a probabilidade posterior da classe (c, alvo) dada preditor (x, atributos).

- $P(c)$ é a probabilidade original da classe.
- $P(x|c)$ é a probabilidade que representa a probabilidade de preditor dada a classe.
- $P(x)$ é a probabilidade original do preditor.

3.5 Coeficiente de Concordância

3.5.1 Acerto Percentual

O Acerto Percentual é calculado através de uma matriz de Confusão, a qual é obtida após o término da execução do método. A sua estrutura é semelhante quando são avaliadas quatro classes de desempenho em que as linhas são as classes de desempenho reais e as colunas as classes de desempenho obtidas. Os resultados presentes na diagonal principal da Matriz de Confusão representam os dados que foram avaliados corretamente, já os demais correspondem aos termos que foram avaliados incorretamente. O acerto percentual é calculado da seguinte maneira:

$$AC = \frac{\sum_{i=1}^M n_{ii}}{\sum_{i=1}^M \sum_{j=1}^M n_{ij}}$$

em que $\sum_{i=1}^M n_{ii}$ é a soma da diagonal principal, $\sum_{i=1}^M \sum_{j=1}^M n_{ij}$ é a soma de todos os elementos da matriz e o M é o total de classes (Moraes; Machado, 2014).

3.5.2 Coeficiente Kappa

O Coeficiente Kappa foi proposto por Cohen (1960), é uma medida robusta ponderada que leva em conta os acertos e erros de acordo com a Matriz de Confusão (Moraes; Machado, 2014). Esse coeficiente é expresso por:

$$K = \frac{(P_0 - P_c)}{(1 - P_c)},$$

onde $P_0 = \sum_{i=1}^M (n_{ii})/N$, $P_c = \sum_{i=1}^M (n_{i+}n_{+i})/N^2$, onde n_{ii} é o total da diagonal principal da Matriz de Confusão; n_{i+} é o total da linha i na Matriz de Confusão; n_{+i} é o total da coluna da mesma matriz; M é o total de classes de desempenho na matriz e N é o número total de decisões presentes na matriz.

Este coeficiente é comumente utilizado para termos de avaliação de treinamento (Duda et al., 2000). A partir dele pode-se distinguir em que situação se encontra a concordância das avaliações em relação a um sistema de referência. De acordo com Landis e Koch (1977), o coeficiente Kappa pode ser interpretado através da Tabela 2.

TABELA 2 – INTERPRETAÇÃO DO COEFICIENTE KAPPA

Coeficiente Kappa	Grau de Concordância
< 0,00	Concordância Pobre
0,00-0,20	Concordância Pequena
0,21-0,40	Concordância Regular
0,42-0,60	Concordância Moderada
0,61-0,80	Concordância Considerável
0,81-1,00	Concordância Quase Perfeita

DESENVOLVIMENTO

As hepatites virais são patologias causadas por diversos agentes etiológicos, de distribuição universal, que têm em comum o hepatotropismo. Possuem similaridades do ponto de vista clínico-laboratorial, mas apresentam marcantes diferenças epidemiológicas e quanto à sua evolução.

Embora apresentem diversas variações de incidência e prevalência, conforme a região geográfica, as hepatites virais representam um problema sanitário de grande relevância, em praticamente todos os países do mundo.

Agrupadas, frequentemente, como doença única, em razão da semelhança de suas apresentações clínicas, elas compreendem entidades mórbidas bem conhecidas e distintas, quanto à etiologia, epidemiologia, evolução, prognóstico e profilaxia. Embora, recentemente, novos vírus tenham sido isolados e, em algum momento, associados a hepatites tem-se como certa, a existência de cinco tipos de hepatites virais, de importância médica (HUANG et al., 2000).

A classificação ocorre com base no seu mecanismo habitual de transmissão, as hepatites virais são comumente agrupadas em dois grandes grupos: o primeiro corresponde àquelas cuja contaminação se dá pelas vias fecal e oral, abrangendo as hepatites A e E; no segundo, situam-se as que se transmitem através de outros mecanismos – predominando a via parenteral – representadas pelas hepatites B, C e Delta (HINRICHSEN et al., 2002).

A relevância das hepatites não se limita ao enorme número de pessoas infectadas; estende-se também às complicações das formas agudas e crônicas. Os vírus causadores das hepatites determinam uma grande variedade de manifestações clínicas, de portador assintomático ou hepatite aguda ou crônica, até cirrose e carcinoma hepatocelular. Considerando que as consequências das infecções são diversas, na dependência do tipo de vírus, o diagnóstico de hepatite, nos dias atuais, será incompleto, a menos que o agente etiológico fique esclarecido (FERREIRA; SILVEIRA, 2004).

Atualmente, a tecnologia e a base de dados de computadores ajudam o homem a coletar e armazenar uma enorme quantidade de dados. O grande tamanho da maioria das bases de dados impossibilita o ser humano de interpretar dados. Portanto, os computadores são necessários para extrair novos A Inteligência Artificial é uma abordagem que pode capacitar os colegas a pensar como seres humanos, onde pode aprender através da experiência, reconhecer padrões de grande quantidade de dados e também o processo de tomada de decisões baseado no conhecimento humano e nas habilidades de raciocínio. Segundo Luger, AI pode ser definido como o ramo da ciência da computação que se preocupa com a automação do comportamento inteligente. É uma combinação de ciência e engenharia para fazer máquinas inteligentes. Existem três perspectivas na IA; 1) AI pode ser como um substituto, 2) pode ser como assistente e 3) também pode ser usado para estender capacidades humanas (KAUR et al., 2012).

Um classificador Naive Bayesiano é um classificador estatístico probabilístico. O termo “ingênuo” refere-se a uma independência condicional entre recursos ou atributos. A suposição “ingênua” reduz complexidade de computação para uma simples multiplicação de probabilidades. Uma das principais vantagens do classificador Naive Bayesian é sua rapidez de uso. Isso porque é o algoritmo mais simples entre algoritmos de classificação (KOTSIANTIS; PINTELAS, 2004). Por causa dessa simplicidade, Ele pode manipular prontamente um conjunto de dados com muitos atributos. Além do que, além do mais, o classificador bayesiano ingênuo precisa apenas de um pequeno conjunto de treinamento dados para desenvolver estimativas de parâmetros precisos porque requer apenas o cálculo das frequências dos atributos e atribuir pares de resultados no conjunto de dados de treinamento. Geralmente, no entanto, o

uso do classificador bayesiano ingênuo produz bom desempenho em termos de precisão de classificação, apesar das violações da suposição de independência de atributos e é, como tal, amplamente utilizado na mineração de dados médicos. Ele também tem sido usado como um algoritmo de linha de base para a comparação de outros tipos de algoritmos de classificação.

RESULTADOS E DISCUSSÃO

As amostras de treinamento e de teste foram geradas a partir da metodologia descrita anteriormente. Na avaliação de teste que seguem a distribuição de Treinamento baseado na Rede Naive Bayes utilizando parâmetros do cross-validation 5 foram encontrados os seguintes resultados para os coeficientes de concordância: utilizando random seed 5 %, temos que o acerto percentual foi de 86,45 % e o coeficiente Kappa obtido foi de 60,04%.

Como observado na tabela 3 esse foi o melhor obtido no treinamento com o cross-validation, acima de 5 tanto acerto percentual quanto coeficiente Kappa diminuíram e se repetem quando é aumentado.

TABELA 3 – TESTE DE TREINAMENTO CROSS VALIDATION

CROSS-VALIDATION	RANDOM SEED %	ACERTO PERCENTUAL	COEFICIENTE KAPPA	CLASSIFICAÇÃO KAPPA
5	5	86.4516 %	60,04 %	Concordância Moderada
10	5	85.8065 %	58,60 %	Concordância Moderada
15	5	85.8065 %	58,60 %	Concordância Moderada
20	5	85.8065 %	58,60 %	Concordância Moderada

Levando em consideração as literaturas o ideal é que o conjunto de dados seja dividido em dois, e geralmente a divisão é de 70% do conjunto para o conjunto de treinamento e 30% para o conjunto de testes. Levando em consideração esses dados o melhor resultado seria acerto percentual foi de 86,95 % e o coeficiente Kappa obtido de 66,50%. Porém o algoritmo naive bayes consegue eficiência com número de instâncias pequenas, mesmo aumentando o banco de dados, ele não teria grandes alterações, por isso, justifica-se o uso do percentil split 90% foram encontrados os seguintes resultados para os coeficientes de concordância: utilizando random

seed 3 %, temos que o acerto percentual foi de 93,33 % e o coeficiente Kappa de 85,71%, melhor resultado obtido para o banco de dados.

TABELA 4 – TESTE DE TREINAMENTO PERCENTIL SPLIT

PERCENTAGE SPLIT	RANDOM SEED %	ACERTO PERCENTUAL	COEFICIENTE KAPPA	CLASSIFICAÇÃO KAPPA
66%	3	86.7925 %	65,62 %	Concordância Moderada
70%	3	86.9565 %	66,50 %	Concordância Considerável
80%	3	90.3226 %	73,95 %	Concordância Considerável
90%	3	93.3333 %	85,71 %	Concordância Quase Perfeita

CONSIDERAÇÕES FINAIS

Este artigo trata dos resultados no campo da classificação de dados obtido com o algoritmo Naive Bayes, um algoritmo fácil e rápido para prever o conjunto de dados da classe de teste. Apesar dos dados ausentes nas instâncias, como também o desbalanceamento dos atributos não comprometeram a eficiência do algoritmo em gerar um modelo para prognóstico da Hepatite. Com o resultado dos experimentos foi possível verificar que o algoritmo tem uma solução satisfatória para classificação dos dados de Hepatite, utilizando o Percentage Split. Em termos de desenvolvimento de método, pode ser usado um método de classificação capaz de processar dados com valor ausente, como árvore de decisão. No futuro, essa pesquisa pode ser desenvolvida tornando-se aplicativos prontos, como aplicativos para Android. Com o aplicativo, em seguida, mais perto do usuário em necessidade.

REFERÊNCIAS

BOBBIO, A.; PORTINALE, L.; MINICHINO, M.; CIANCAMERLA, E.. Improving the Analysis of Dependable Systems by Mapping Fault Trees into Bayesian Networks. *Reliability Engineering & System Safety*, Vol. 71, p.249-260, 2001.

COHEN, J. A coefficient of agreement for nominal scales. In: *Educat. Psyc. Measurement*. [S.l.: s.n.], 1960.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2nd. ed. [S.l.]: Wiley Interscience, 2000.

FERREIRA, Cristina Targa; SILVEIRA, Themis Reverbel da. Hepatites virais: aspectos da epidemiologia e da prevenção. **Revista Brasileira de Epidemiologia**, v. 7, p. 473-487, 2004.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. Morgan Kaufmann, 2001.

HAN, J. and KAMBER, M. (2006). *Data Mining: Concepts and Techniques*. **Morgan Kaufmann Publishers**, 2nd ed.

HINRICHSEN, H. et al. Prevalence and risk factors of hepatitis C virus infection in haemodialysis patients: a multicentre study in 2796 patients. **Gut**, v. 51, n. 3, p. 429-433, 2002.

HUANG, Yi-Hsiang et al. Prevalence and risk factor analysis of TTV infection in prostitutes. **Journal of medical virology**, v. 60, n. 4, p. 393-395, 2000.

JAYAKUMAR S, Chowdhury R, Ye C, Karvellas CJ. **Fulminant viral hepatitis**. *Crit Care Clin*. 2013;29(3):677-97

KAUR, Harpreet et al. Artificial Intelligence: Bringing expert knowledge to computers. **Discov. J**, v. 2, p. 4-7, 2012.

KOTSIANTIS, Sotiris B.; PINTELAS, Panayiotis E. Increasing the classification accuracy of simple bayesian classifier. In: **International Conference on Artificial Intelligence: Methodology, Systems, and Applications**. Springer, Berlin, Heidelberg, 2004. p. 198-207.

KRAUSE P.J., "Learning Probabilistic Networks," **Philips research laboratories**, 1998.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. In: *Biometrics*. [S.l.: s.n.], 1977. p. 159-174

MITCHELL, T. (1997). **Machine Learning**. McGraw Hill.

MORAES, R. M.; MACHADO, L. S. Psychomotor skills assessment in medical training based on virtual reality using a weighted possibilistic approach. In: *Knowledge Based Systems*. [S.l.: s.n.], 2014. p. 97-62.

PASSAIA, Nereu Adilar; PILLA, Bianca Smith. A tecnologia da informação aplicada à tomada de decisão em consultórios médicos. **Estudos do CEPE**, v. 65, n. 32, p. 65-93, 2010.

PODDAR B, SAIGAL S, KUMAR A, SINGH RK, AZIM A, GURJAR M, et al. Factors associated with outcome in acute liver failure in an intensive care unit. *Indian J Gastroenterol*. 2013;32(3):172-8.

RAJESH, R.; MAITI, J.; REENA, M. Decision Tree for Manual Material Handling Tasks Using WEKA. In: **Ergonomic Design of Products and Worksystems-21st Century Perspectives of Asia**. Springer, Singapore, 2018. p. 13-24.

SANTIKA I. W., “Pengembangan Sistem Pakar Konsultasi Hama Dan Penyakit Tanaman Jeruk Menggunakan Metode Bayesian Network Berbasis Web,” Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika, Vol. 1, No. 4, 2012.

UNIVERSITY OF WAIKATO. **Weka 3 – Machine Learning Software in Java**. Disponível no site da University of Waikato (2010).

VIEIRA FILHO, V.; ALBUQUERQUE, M. T. C. F. . Abordagem Bayesiana para Simulação de Jogos Complexos. In: SBGames, 2007, São Paulo. Proceedings of SBGames 2007, 2007.

WITTEN, I.; FRANK, E. **Data Mining – Practical Machine Learning Tools**. Morgan Kaufmann, 2000.

WEDEMEYER, H; MANNS, M. P. Epidemiology, pathogenesis and management of hepatitis D: update and challenges ahead. Nature Reviews Gastroenterology and Hepatology, London, v. 12, n. 4, p. 31-40, 2015. Disponível em: <http://www.nature.com/nrgastro/journal/v7/n1/abs/nrgastro.2009.205.html>. Acesso em: 05 Dez. 2015.