

# MINERAÇÃO DE DADOS PARA AVALIAÇÃO DOS FATORES DE RISCO PARA A DOENÇA RENAL CRÔNICA

Lilian Mie Mukai Cintho (PUCPR) miemukai@hotmail.com  
Deborah Ribeiro Carvalho (PUCPR) ribeiro.carvalho@pucpr.br  
Marlene Harger Zimmermann (PUCPR e UEPG) 061040449@uepg.br

**Resumo:** A doença renal crônica tem sua incidência e prevalência em constante elevação no Brasil. Em 2012 ocorreu o Dia Mundial do Rim, em Ponta Grossa (PR), quando foram coletados dados de 283 participantes envolvendo 12 variáveis das quais foram extraídas estatísticas e padrões que permitissem identificar relações entre os fatores de risco. Para a extração de padrões foram utilizados algoritmos de mineração de dados. Verificou-se que a maioria dos participantes desconhece ser portador de hipertensão arterial e diabetes mellitus, sendo que 52,3% possuem risco substancialmente aumentado para doenças cardiovasculares. Outra situação evidenciada é a interferência do grau de escolaridade na prática de atividade física. Também a partir dos resultados obtidos pela mineração de dados foi possível perceber a associação entre sobrepeso e faixa de glicemia entre 70 e 140 mg/dl, porém quando o sobrepeso estava associado a indivíduos de cor parda esta faixa de glicemia se altera para 140 a 200 mg/dl. A exceção determinada pelo fato do indivíduo pertencer ao grupo cor parda constitui um fator de risco importante para o diagnóstico da hipertensão arterial e diabetes mellitus. Outros resultados obtidos a partir da mineração de dados demonstraram que o valor de corte identificado para glicemia, para caracterizar diabetes mellitus ou não diabetes mellitus, foi 162 mg/dl, quando o inicialmente estabelecido pelos protocolos da Sociedade Brasileira de Diabetes é 100 mg/dl. A aplicação da mineração de dados oportunizou complementar o conhecimento do profissional de saúde para a identificação dos fatores de risco da população desenvolver a doença renal crônica.

**Palavras-chave:** Mineração de Dados; Insuficiência Renal Crônica; Fatores de Risco.

## DATA MINING FOR ASSESSMENT OF RISK FACTORS FOR CHRONIC KIDNEY

**Abstract:** Chronic kidney disease has its incidence and prevalence in constantly increasing in Brazil. In 2012, the World Kidney Day was promoted in Ponta Grossa (PR), which in data was collected from 283 participants comprising 12 variables of which were extracted statistics and patterns that allow identifying relationships between risk factors. For extracting patterns were used for data mining algorithms. It was found that the majority of participants being unaware carrier hypertension and diabetes mellitus, of which 52.3% have a substantially increased risk for cardiovascular disease. Another situation is evidenced interference schooling in physical activity. Also according of obtained results by data mining was possible to connect the association between overweight and glucose range between 70 and 140 mg/dl, but when overweight was associated with individuals of mixed race this track glucose changes from 140 to 200 mg/dl. An important risk factor for the diagnosis of hypertension and diabetes mellitus is the exception determined by whether the individual belongs to the group mulatto. Other results detected from the data mining showed that the cutoff value identified for glucose, to characterize diabetes mellitus or no diabetes mellitus was 162 mg / dl, when initially established by the protocols of the Brazilian Diabetes Society is 100 mg / dl. The application of data mining provided an opportunity to complement the knowledge of health professionals to identify the risk factors population develop chronic kidney disease.

**Keywords:** Data mining; Chronic Kidney Disease; Risk factors.

## 1. INTRODUÇÃO

A doença renal crônica (DRC) é considerada um problema de Saúde Pública Mundial (BASTOS; BREGMAN; KIRSZTAJN, 2010). Levantamento realizado por Nogueira et al. (2011), na Nigéria, foram identificados 2,5 casos por milhão de portadores de DRC e nos EUA, cerca de 13,1% da população é afetada pela doença (BASTOS; KIRSZTAJN, 2011). No Brasil, segundo dados do Ministério da Saúde (BRASIL, 2012), no período de 2011, foram gastos pelo Sistema Único de Saúde (SUS) cerca de 1,8 bilhões em hemodiálise, com

registro de 15 mil mortes por consequência da patologia renal. Segundo o censo da Sociedade Brasileira de Nefrologia (SBN), em 2013, estima-se que 100 mil pacientes encontram-se em Terapia Renal Substitutiva (TRS), número este que vem crescendo anualmente de forma preocupante (SOCIEDADE BRASILEIRA DE NEFROLOGIA, 2013).

As principais causas da DRC são a hipertensão arterial (HA) e diabetes mellitus (DM) que correspondem a 35% e 30% respectivamente. A DRC é considerada doença crônica não transmissível, assim como a HA e DM e são passíveis de controle para evitar a sua progressão (SOCIEDADE BRASILEIRA DE NEFROLOGIA, 2013).

É considerado portador de DRC, todo indivíduo adulto, com idade acima de 17 anos, por um período maior que três meses, apresentar filtração glomerular (FG) menor que 60 ml/min/1,73m<sup>2</sup> ou FG maior que 60 ml/min/1,73m<sup>2</sup> com pelo menos um marcador de lesão do parênquima renal como a proteinúria e/ou hematuria glomerular ou alteração no exame de imagem (raios x simples de abdome, ultrassonografia dos rins e vias urinárias ou tomografia) presente por um período igual ou superior a três meses (K/DOQI, 2002).

Pertencem ao grupo de risco da DRC os indivíduos portadores de HA, DM, familiares com história de DRC, pacientes em uso de medicamentos nefrotóxicos, obesos, idosos e portadores de doença cardiovascular (DCV) (ROMÃO JR, 2004; BASTOS, BREGMAN; KIRSZTAJN, 2010; CANZIANI, 2004).

Considera-se hipertenso o indivíduo com idade acima de 18 anos, com pressão arterial sistólica maior ou igual a 140 mmHg ou pressão diastólica maior ou igual a 90 mmHg, causada por condição clínica multifatorial. A HA frequentemente altera as funções e ou estruturas dos órgãos-alvo como coração, encéfalo, rins e vasos sanguíneos e também o metabolismo que traz o aumento para risco de eventos cardiovasculares (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2010).

Define-se como DM “um grupo de doenças metabólicas, caracterizada por hiperglicemia e associada a complicações, disfunções e insuficiência de vários órgãos especialmente olhos, rins, nervos, cérebro, coração e vasos sanguíneos” (BRASIL, 2006).

Programas de promoção à saúde, diagnóstico e encaminhamento precoce ao nefrologista para tratamento com controle da pressão arterial, glicemia, restrição da ingesta proteica e lipêmica, da anemia, mudança de estilo de vida com redução de peso, atividade física regular, do fumo, do álcool e adesão ao tratamento medicamentoso promovem o controle da progressão da DRC e DCV (LEVIN et al., 2008).

Os serviços de saúde têm como desafios realizar o rastreamento dos indivíduos com fatores de risco (FR) para doenças crônicas não transmissíveis e de desenvolver programas de prevenção à saúde para prevenção e controle do DM e HA (TOSCANO, 2004) que são FR para a DRC.

No Brasil, o governo implantou o programa Hiperdia (Sistema de Cadastramento e Acompanhamento de Hipertensos) para auxiliar na prevenção e progressão quanto ao aumento de novos casos de HA e DM e realizou campanha Nacional para a respectiva detecção (BRASIL, 2012; TOSCANO, 2004). A Sociedade Internacional de Nefrologia criou o Dia Mundial do Rim, em oito de março no qual a SBN é a responsável nacional pela promoção de ações preventivas para a educação sobre a saúde renal (SOCIEDADE BRASILEIRA DE NEFROLOGIA, 2010). Além desses programas, várias pesquisas são realizadas mundialmente para rastreamento de indivíduos portadores de FR para o desenvolvimento de HA, DM (TOSCANO, 2004) e da DRC (SOCIEDADE BRASILEIRA DE NEFROLOGIA, 2010).

Em conformidade com este conjunto de ações educativas ocorreu uma ação conduzida pelos discentes e docentes do Curso de Enfermagem, farmácia e medicina da Universidade Estadual de Ponta Grossa (PR), ou seja, o Dia Mundial do Rim, num supermercado em Ponta Grossa, em 2012.

O rastreamento da população de maior risco para o desenvolvimento da DRC é relevante para prevenir que a doença se instale ou mesmo para realizar o diagnóstico precoce e encaminhamento ao especialista para o controle da evolução da doença. Embora haja controvérsias sobre qual seja a melhor forma de rastreamento, é recomendado realizá-la de forma seletiva, envolvendo indivíduos pertencentes ao grupo de risco como exemplo, a HA, DM, idosos e familiares de DRC (SOCIEDADE BRASILEIRA DE NEFROLOGIA, 2010).

As diretrizes clínicas para estratificar os indivíduos sob risco de desenvolver a DRC, são claras, porém, será que existe alguma relação importante entre os FR que influencie no diagnóstico da DRC ?

Atualmente, com a possibilidade de armazenar grandes quantidades de dados a partir dos sistemas computacionais, surgem novas alternativas tecnológicas para a sua melhor utilização, entre elas a Mineração de Dados (CAMILO; SILVA, 2009).

A descoberta de conhecimento em bases de dados, Knowledge Discovery in Data Bases (KDD) é um processo que permite potencializar a utilização dos dados, sendo a Mineração de Dados uma das suas respectivas etapas. Mineração de Dados é definida como “o processo não trivial de identificação de padrões válidos, desconhecidos, potencialmente úteis e, no final das contas, compreensíveis em dados” (FAYYAD; SHAPIRO; SMYTH, 1996). Para Han e Kamber (HAN; KAMBER, 2000) Mineração de Dados refere-se à extração de conhecimento de grande quantidade de dados armazenados tanto em banco de dados como em data warehouses ou outros repositórios. Tem como função descrever classes e/ou conceitos, realizar análise associativa, classificação e predição, análise de agrupamento, análise de outliers e análise de evolução de dados.

O processo de KDD é interativo que envolve quatro etapas (FAYYAD; SHAPIRO; SMYTH, 1996):

- Compreensão do domínio da apuração
- Preparação do conjunto de dados sobre o qual a descoberta será realizada, a partir das diversas fontes disponíveis;
- Aplicação de algoritmos de Mineração de Dados; e
- Análise dos padrões descobertos

A Mineração de Dados atende as seguintes tarefas: descrição, classificação, estimação ou regressão, predição, agrupamento e descoberta de regras de associação (CAMILO; SILVA, 2009).

Uma regra de associação permite identificar quais itens de dados estão relacionados entre si, no seguinte formato: se X então Y, no qual X e Y são conjuntos de itens de dados, sendo X considerado antecedente da regra e Y o seu consequente (CAMILO, SILVA, 2009).

Na classificação os padrões descobertos são de tal forma representados que permitem predizer a qual classe um novo registro, ainda não classificado, pertence. Existem diversas formas para representar um classificador, uma das mais usuais é a árvore de decisão (CARVALHO et al., 2012).

Muitas vezes o conjunto de padrões descoberto excede a capacidade de análise e interpretação do ser humano, exigindo a adoção de estratégias de pós-processamento que facilitem esta etapa. Entre as diversas estratégias se destaca a descoberta de pares de regras que representem a regra geral e a sua respectiva regra de exceção.

Exemplo de regra geral R1: Se X então Y

Exemplo de regra de exceção R2: Se X e Z então  $\neg$  Y

A regra de exceção (R2) é uma especialização de uma regra geral (R1) e altera o consequente previsto pela regra geral (de Y para não ( $\neg$ ) Y, na medida que o antecedente Y foi combinado ao Z. Em geral as regras de exceção tendem a ser mais interessantes para o usuário, por representarem uma contradição em relação ao conhecimento geral já disponível ao usuário (HUSSAIN et al., 2000).

A partir dos dados coletados, durante o Dia Mundial do Rim, este trabalho teve como objetivo utilizar a mineração de dados para a descoberta de relações entre os FR possibilitando um rastreamento ainda mais adequado dos participantes com o objetivo de para apoiar a identificação da população com maior risco para o desenvolvimento da DRC.

## **2. MATERIAIS E MÉTODOS**

A população em estudo foram os participantes do Dia Mundial do Rim, que ocorreu em 31 de março de 2012, no período das 08h00min às 18h30min, em um supermercado de Ponta Grossa (PR). Ao todo 283 adultos aceitaram participar de forma voluntária do evento. Os dados foram coletados a partir de questionário contendo 37 perguntas fechadas, elaboradas de acordo com a proposta de uniformização da coleta de dados da SBN (KIRSZTAJN; BASTOS, 2007). As variáveis coletadas foram: gênero, cor, idade, escolaridade, história clínica, histórico familiar, tabagismo e a realização ou não de atividade física.

Além de dados sobre o perfil foram medidas pressão arterial, peso, altura, circunferência abdominal (CA) e a glicemia capilar. Do total de variáveis coletadas, foram selecionadas gênero, cor, idade, escolaridade, diagnóstico para DM e HA, histórico de tabagismo, atividade física e de problemas renais; peso, altura e circunferência abdominal.

O critério que orientou a seleção destas variáveis foi a aderência ao objetivo deste experimento que é verificar o que interfere no desenvolvimento da DRC em relação aos hábitos de vida, conhecimento prévio dos FR desencadeantes da DRC e se a cor, gênero e escolaridade interferem no desenvolvimento da DRC. Foram excluídos para este estudo os dados referentes àqueles que sabiam ser portadores de HA e DM e que realizavam tratamento e controle destas patologias, uma vez já possuíam acompanhamento e orientação médica.

O conjunto de dados disponível em uma planilha eletrônica, foi pré-processado transformando os valores numéricos em valores categóricos de acordo com intervalos preconizados pela literatura (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2010; SOCIEDADE BRASILEIRA DE DIABETES, 2011; SOCIEDADE BRASILEIRA DE ENDOCRINOLOGIA E METABOLOGIA, 2004). Como exemplo, destaca-se a circunferência abdominal (CA), glicemia (Gl), pressão arterial e índice de massa corporal (IMC).

Para transformar a medida da CA em valores nominais foi adotado o critério preconizado por Lean e cols. (SOCIEDADE BRASILEIRA DE ENDOCRINOLOGIA E METABOLOGIA, 2004) de acordo com o grau de risco para doenças cardiovasculares. Os valores referem-se a homens e mulheres caucasianos no qual é considerado CA normal em homem com até 94 cm e na mulher até 80 cm. Risco aumentado em homem quando CA for maior ou igual a 94 cm e na mulher CA maior ou igual a 80 cm. Risco aumentado substancialmente no homem quando CA for maior ou igual a 102 cm e na mulher, CA maior ou igual a 88 cm.

Para a pressão arterial foi adotada a diretriz da Sociedade Brasileira de Cardiologia (2010) sendo atribuído no estudo os valores da pressão normal conforme classificação apresentada na Tabela 1 e considerado hipertenso, valores da pressão sistólica entre 140 a 159 mmHg e pressão diastólica entre 90 a 99 mmHg.

**Tabela 1**– Classificação da pressão arterial, em adulto, maior ou igual a 18 anos

Classificação da pressão arterial	Pressão sistólica mmHg	Pressão diastólica mmHg
Ótima	< 120	< 80
Normal	< 130	< 85
Limítrofe	130-139	85-89
Hipertensão estágio 1	140-159	90-99
Hipertensão estágio 2	160-179	100-109
Hipertensão estágio 3	≥ 180	≥ 110
Hipertensão sistólica isolada	≥ 140	< 90

Fonte: Sociedade Brasileira de Cardiologia, 2010.

O diagnóstico da DM segue o critério aceito pela Organização Mundial da Saúde (OMS) e pela Sociedade Brasileira de Diabetes (2011), apresentada na Tabela 2. A hiperglicemia é o aumento da glicose no sangue em que é considerada alterada quando os valores da GI:

- em jejum estiverem acima de 126 mg/dl;
- maior ou igual a 200 mg/dl, duas horas após a carga de 75g de glicose; e
- maior ou igual a 200 mg/dl ao acaso.

**Tabela 2**– Classificação Glicemia. Fonte: Sociedade Brasileira de diabetes, 2012-2013

Categoria	Jejum	Casual	2 horas após sobrecarga com 75 g de glicose
GI normal	< 100 mg/dl		
Diabetes Mellitus	≥ 126 mg/dl	≥ 200 mg/dl (com sintomas clássicos)	≥ 200 mg/dl

Fonte: Sociedade Brasileira de diabetes, 2012-2013.

A Tabela 3 apresenta a classificação do IMC adaptada pela World Health Organization baseada em padrões internacionais para indivíduos com 18 anos ou mais (SOCIEDADE BRASILEIRA DE ENDOCRINOLOGIA E METABOLOGIA, 2004).

**Tabela 3**– Classificação de peso pelo IMC

Classificação	IMC (Kg/m <sup>2</sup> )
Baixo peso	< 18,5
Normal	18,5 a 24,9
Sobrepeso	25 a 29,9
Obesidade grau I	30 a 34,9
Obesidade grau II	35 a 39,9
Obesidade Grau III (mórbido)	>40

Fonte: Sociedade Brasileira de endocrinologia e metabologia, 2004.

Para caracterizar o perfil dos pacientes analisados, foram extraídas algumas frequências envolvendo as variáveis dos critérios de inclusão utilizando a ferramenta SPSS - Statistical Package for Social Sciences (MUNDSTOCK et al., 2006).

Além das frequências também foram realizados dois experimentos envolvendo Mineração de Dados: descoberta de regras de associação e descoberta de classificador na forma de árvore de decisão. Para o primeiro experimento foi utilizado o algoritmo APRIORI (BORGELT, 2004), sendo que sobre o conjunto descoberto identificadas regras gerais e as suas respectivas exceções (GUIMARÃES; MORAES; CARVALHO, 2006), (MILANI; CARVALHO, 2013).

Para a descoberta do classificador na forma de árvore de decisão foi adotado o algoritmo J48 (STEPHEN, 2010), considerando os parâmetros default e posteriormente

transformada a árvore em um conjunto de regras do tipo “se”... (condições), “então”... (classe) (CARVALHO et al., 2012). Foram descobertas duas árvores diferindo o atributo classe, a saber: o status em relação à DM e o status em relação à HA.

A partir da análise do conjunto de regras oriundas da transformação das árvores foram identificadas algumas situações de interesse que motivaram a retornar a base de dados original para a extração de novas estatísticas. Esta motivação esteve mais fortemente relacionada a situações de “erros de classificação”, como verificado na associação entre a o IMC de sobrepeso e Gl entre 70 a 140 mg/dl. Observou-se que esta condição é contrária quando o IMC for sobrepeso mas os indivíduos forem de cor parda elevando a Gl para a faixa de 140 a 200 mg/dl.

Os dados obtidos foram resultados das ações propostas pelo Projeto de Extensão de Prevenção da Doença Renal Crônica da Universidade Estadual de Ponta Grossa (UEPG) do Curso de Enfermagem sem a identificação do nome dos participantes e que de acordo com os artigos 9, 10 e 11 do capítulo 9 da tramitação e aprovação da resolução do CEPE nº 235, da UEPG de 8 de dezembro de 2009, não exige submissão ao comitê de ética de projeto específico para a pesquisa.

### 3. RESULTADOS E DISCUSSÕES

Reconhecer quem são os indivíduos portadores de FR para desenvolver a DRC é importante para o diagnóstico precoce da doença. Desta forma foram discutidos neste trabalho os principais FR associados à causa da DRC pelo perfil encontrado na análise estatística descritiva e a partir da identificação da mineração de dados.

Como perfil encontrado na população em estudo, à partir da estatística descritiva foi possível perceber que dos valores pressóricos, 209 (73,8%) indivíduos apresentaram pressão arterial dentro dos parâmetros de normalidade, entretanto, dos 39 (13,8%) indivíduos que apresentaram pressão arterial acima dos valores considerados normais, apenas 25 (64%) destes sabiam ser portadores da HA.

A HA é considerada uma doença silenciosa, assintomática e que por isso é necessário esforços concentrados dos profissionais da saúde, sociedades científicas e das agências governamentais para se atingir metas efetivas de diagnóstico, tratamento e controle da hipertensão (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2010) uma vez que é uma das causas da DRC. Estudos epidemiológicos brasileiros comprovam que uma parcela importante da população adulta com HA não sabe ser portador de HA (TOSCANO, 2004) o que vem de encontro com o resultado deste trabalho e como consequência, a falta do diagnóstico precoce da DRC.

Quanto aos índices glicêmicos, 9 participantes (3,2%) tiveram valor glicêmico superior a 200mg/dl e desses, 7 (78%) sabiam ser portadores de DM. Estudo realizado pelo Ministério da Saúde, Sociedade Brasileira de Diabetes e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), em 1987, demonstraram que cerca de 50% da população não sabiam ser portadores de DM o que vem a demonstrar que muitos deles encontram-se sem diagnóstico e correm o risco de desenvolverem complicações decorrentes da patologia como a DRC (BARBOSA; BARCELÓ; MACHADO, 2001).

Em relação à CA, 52,3% apresentaram risco aumentado substancialmente para DCV e destes, 73% eram do sexo feminino. Dos 24,7% dos portadores de HA, 77,8% apresentaram CA com risco aumentado substancialmente para DCV e 10,2% dos portadores de DM, 75,9% possuíam risco aumentado substancialmente para DCV. Na Universidade Federal de Viçosa (SAVIO et al., 2008), foi realizado um estudo e constatado percentual de CA elevado em mulheres (42%) e homens (22%) para risco aumentado substancialmente para DCV e que a gordura abdominal eleva a Gl e a pressão arterial corroborando com os dados encontrados

neste estudo. A DCV é a principal causa de morbimortalidade na população com DRC, principalmente àqueles em TRS (K/DOQI, 2005).

Dos 5,7% dos participantes que possuíam 3º grau incompleto de escolaridade, 81,3% não praticavam atividade física. Em comparação com um estudo (SAVIO et al., 2008), observou-se que indivíduos com grau de escolaridade superior praticavam menos atividade física em relação aos indivíduos que possuíam apenas o primeiro grau, confirmando o resultado encontrado de 81,3% dos participantes de 3º grau incompleto não praticarem atividade física. A Sociedade Brasileira de Cardiologia (2010) relata que a atividade física reduz a incidência de HA bem como a mortalidade e o risco cardiovascular e conseqüentemente ao risco de desenvolver a DRC.

As estratégias para estratificar pessoas sob o risco de desenvolver a DRC é fundamental para o diagnóstico precoce e para isso, é necessário prevenir, tratar e controlar os FR que causam a DRC.

A partir da mineração de dados foram descobertas 8.792 regras, sendo identificados pares envolvendo regras gerais e as respectivas regras de exceção tais como:

**Regra Geral1:**

Se o IMC for sobrepeso então a Gl se encontra entre 70 e 140 mg/dl (37,5%, 74,5%).

Lê-se: 37,5% possuem IMC de sobrepeso e desses, 74,5% apresentam Gl entre 70 a 140 mg/dl.

**Regra de Exceção 1:**

Se o IMC for sobrepeso e de cor parda então a Gl se encontra entre 140 e 200 mg/dl (0,4%, 100,0%).

Lê-se: 0,4% possuem IMC de sobrepeso e cor parda e desses, 100% apresentam Gl entre 140 a 200 mg/dl

**Regra Geral 2:**

Se a idade for entre 61 a 70 anos então Gl se encontra entre 70 e 140 mg/dl (7,1%, 80%).

Lê-se: 7,1% dos participantes possuem idade entre 61 a 70 anos, e desses, 80% apresentam Gl entre 70 a 140 mg/dl.

**Regra de Exceção 2:**

Se idade for entre 61 a 70 anos e de cor parda então Gl se encontra entre 140 a 200 mg/dl (0,4%, 100%).

Lê-se: De 0,4% dos participantes com idade entre 61 a 70 anos e cor parda, 100% apresentam Gl entre 140 a 200 mg/dl.

**Regra Geral 3:**

Se IMC for sobrepeso então não se encontra hipertenso (37,5%, 77,4%).

Lê-se: De 37,5% dos participantes possuem IMC de sobrepeso e desses, 77,4% não são hipertensos.

**Regra de Exceção 3:**

Se IMC for sobrepeso e de cor parda então se encontra hipertenso (2,8%, 75%).

Lê-se: àqueles que apresentam IMC de sobrepeso e cor parda (2,8%), 75% são hipertensos.

**Regra Geral 4:**

Se CA for de risco aumentado para doença cardiovascular então se encontra não hipertenso (27,2%, 88,3%).

Lê-se: de 27,2% dos participantes possuem CA de risco aumentado para DCV e desses, 88,3% não são hipertensos.

Regra de Exceção 4:

Se CA for de risco aumentado para doença cardiovascular e de cor parda então se encontra hipertenso (1,4%, 75%).

Lê-se: 1,4% dos participantes possuem CA de risco aumentado para doença cardiovascular e cor parda, desses, 75% são hipertensos.

Vale destacar que sobre o conjunto total de regras descobertas (8.792 regras) foram identificados 332 subconjuntos de regras gerais e suas respectivas exceções. Situação de destaque, relacionado a FR, envolvem pessoas de cor parda, uma vez que está fortemente associada a presença do DM ou HA.

A relação da cor ainda não está bem elucidada, porém estudos evidenciam indivíduos afrodescendentes com maior prevalência à DM e HA (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2010; SOCIEDADE BRASILEIRA DE DIABETES, 2012-2013). Assim, verificou-se que com a regra de exceção, o atributo cor parda alterou a regra de decisão, evidenciando a relevância em buscar estas regras de exceções para a detecção precoce da DRC e determinar um rastreamento seletivo aos portadores de DM e HA (BASTOS R.; BASTOS M; TEIXEIRA, 2007).

A árvore de decisão descoberta pelo algoritmo J48 (STEPHEN, 2010), considerando o status de DM como classe predita, apresenta apenas duas ramificações (regra 5 e 5.1), envolvendo o valor coletado para a Gl, ou seja, se o valor for  $\leq 162$  então não DM, caso contrário então DM.

Regra 5: se  $Gl \leq 162$  então não DM (265/ 17.0)

Regra 5.1: se  $Gl > 162$ : DM (17/5)

A regra 5 indica que 265 pessoas do conjunto tem  $Gl \leq 162$ mg/dl, destas 265 pessoas 17 (6,4%) são DM, ou seja contrariam o valor da classe predito pela regra (então não DM), constituindo assim um erro de predição. Já na regra 5.1, 17 participantes apresentaram  $Gl > 162$ mg/dl, tendo como classe predita pela regra DM, porém 5 participantes (29%) eram originalmente não DM, mas que na verdade não sabiam ser ou não ser DM.

Nota-se a importância do controle rigoroso da Gl em portadores de DM com a finalidade de intervir na progressão da patologia e no desenvolvimento da DRC (BASTOS; BREGMAN; KIRSZTAJN, 2010) uma vez que a maioria dos participantes que possuem o diagnóstico de DM, não tem o controle da Gl.

Considerando o status DM pela classificação da árvore de decisão, mostra que pelo fato dos participantes não estarem em jejum de oito horas para a coleta de Gl, o algoritmo alterou os valores da Gl para 162 mg/dl denotando a importância de se verificar a Gl conforme os protocolos da Sociedade Brasileira de Diabetes (2012-2013).

Ao construir o classificador considerando a variável status da HA como sendo a classe predita, a árvore de decisão descoberta contempla 10 regras predizendo os valores de status pressóricos como segue:

Para pressão arterial = normotensos

Regra 6:  $Gl \leq 149$ : não hipertenso (195/25)

A regra 6 indica que entre os participantes ditos normotensos, 195 apresentaram GI  $\leq 149$ mg/dl, tendo como classe predita não hipertensos. Porém destes 195 que apresentavam este nível glicêmico, 25 participantes eram hipertensos, ou seja contrariavam a classe predita, representando uma taxa de erro de 12,8%.

Para pressão arterial = hipertensos  
Regra 7: cor = branca  
altura  $\leq 172$  cm: hipertenso (22.28/5.14)

A regra 7 indica que entre os participantes ditos hipertensos, 22 pessoas do conjunto têm cor branca e altura  $\leq 172$  cm, tendo como classe predita hipertenso. Porém destes 22 que apresentavam cor branca e altura  $\leq 172$  cm, 5 participantes eram não hipertensos, ou seja contrariavam a classe predita, representando uma taxa de erro de 22,7%.

Pressão arterial = limítrofe  
Regra 8: problema renal = teve problema renal: hipertenso (6.0/1.0)  
Regra 8.1: problema renal = não teve problema renal: não é hipertenso (24.32/4.21)

A regra 8 indica que entre os participantes ditos limítrofes, 6 pessoas apresentaram problema renal, tendo como classe predita hipertenso. Porém destas 6 que apresentavam hipertensão, ou seja contrariavam a classe predita, representando uma taxa de erro de 16,6%.

A regra 8.1 indica que entre os participantes ditos limítrofes, 24 pessoas não apresentaram problema renal, tendo como classe predita não hipertenso. Porém destas 4 que apresentaram como sendo hipertensas, ou seja contrariavam a classe predita, representando uma taxa de erro de 16,6%.

Para pressão arterial = hipotenso  
Regra 9: pressão arterial= hipotenso: não hipertenso (2/0)

A regra 9 indica que entre os participantes ditos hipotensos, 2 pessoas apresentaram hipotensão, tendo como classe predita não hipertenso.

Em relação à árvore de decisão para a HA, surpreende o fato do algoritmo ter selecionado o atributo GI para caracterizar os não HA tendo em vista que estavam também disponíveis para o algoritmo atributos como peso, IMC e baixo nível de atividade física que em geral estão fortemente relacionados com o status da HA, no qual estima-se que 60% dos HA apresentam mais de 20% de sobrepeso (KOTCHEN, 2008).

Em relação à cor, a árvore de decisão identificou a cor branca como sendo o classificador para a hipertensão, porém, verifica-se que a hipertensão é duas vezes mais prevalente em indivíduos negros o que vem a contradizer o resultado deste estudo (SOCIEDADE BRASILEIRA DE CARDIOLOGIA, 2010). E por último, os problemas renais estão fortemente relacionados como sendo uma das complicações da HA o que vem a corroborar com o resultado encontrado na árvore de decisão dos participantes.

#### 4. CONCLUSÕES

O resultado deste trabalho demonstra os FR presentes em uma população para o desenvolvimento da DRC. Verifica-se a necessidade de intervenções aos FR encontrados com ganho na utilização da tecnologia de mineração de dados pela associação de valores que auxiliam a compreender as necessidades de mudanças no estilo de vida e da capacitação do indivíduo ao autocuidado para prevenir e tratar futuros problemas que acarretarão à sua saúde,

bem como levantar um indicador relevante no rastreamento e diagnóstico precoce da HA e DM para intervir na progressão da DRC.

As etapas de classificação, descoberta de regras de associação dos dados oportunizou melhor entendimento das especificidades dos participantes do evento oportunizando melhor conhecimento ao profissional de saúde à identificação da população com maior risco para desenvolver a DRC bem como as condutas a serem adotadas.

Como pesquisas futuras se recomenda que novos experimentos com a mineração de dados sejam realizados ampliando o conjunto de variáveis permitindo assim que os programas descubram outros padrões que possam vir a ser interessantes na identificação de FR.

## AGRADECIMENTOS

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo.

## REFERÊNCIAS

BARBOSA, R.B.; BARCELÓ, A.; MACHADO, C.A. *Campanha nacional de detecção de casos suspeitos de diabetes mellitus no Brasil: relatório preliminar 1. Pan American Journal of Public Health*, v.10, n.5, nov, 2001.

BASTOS, R.M.R.; BASTOS, M.G.; TEIXEIRA, M.T.B. *A Doença renal crônica e os desafios da atenção primária à saúde na sua detecção precoce. Revista APS*, v.10, n.1, 2007.

BASTOS, M.G.; BREGMAN, R.; KIRSZTAJN, G.M. *Doença renal crônica: frequente e grave, mas também prevenível e tratável. Rev Assoc Med Bras*, v.56, n.2, p.248-253, 2010.

BASTOS, M.G.; KIRSZTAJN, G.M. *Doença renal crônica: importância do diagnóstico precoce, encaminhamento imediato e abordagem interdisciplinar estruturada para melhora do desfecho em pacientes ainda não submetidos à diálise. J Bras Nefrol*, v.33, n.1, p.93-108, 2011.

BORGELT, C. *Apriori: Association rule induction*, 2004. Disponível em <http://www.borgelt.net/apriori.html>. Acesso em 15/01/2013.

BRASIL. MINISTÉRIO DA SAÚDE. *Caderno de atenção básica: Diabetes Mellitus, nº 16. Brasília*, 2006. Disponível em [http://bvsmms.saude.gov.br/bvs/publicacoes/diabetes\\_mellitus.PDF](http://bvsmms.saude.gov.br/bvs/publicacoes/diabetes_mellitus.PDF). Acesso em 15/08/2012.

BRASIL. MINISTÉRIO DA SAÚDE. *Dia mundial do rim. Recursos para hemodiálise aumentam em R\$ 182,6 milhões. Disponível em* [http://portalsaude.saude.gov.br/portalsaude/noticia/4458/162/recursos-para-hemodialise-aumentam-em-r\\$-1816-mi.html](http://portalsaude.saude.gov.br/portalsaude/noticia/4458/162/recursos-para-hemodialise-aumentam-em-r$-1816-mi.html). Acesso em 12/08/2012.

CAMILO, C.A.; SILVA, J.C. *Mineração de dados: conceito, tarefas, métodos e ferramentas. Instituto de Informática. Universidade Federal de Goiás*, 2009. Disponível em [http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-09.pdf](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf). Acesso em 11/11/2012.

CANZIANI, M.E.F. *Diretriz da Sociedade Brasileira de Nefrologia. Doenças cardiovasculares na doença renal crônica. J Bras Nefrol*, v.26, n.3, p.11-14, ago, 2004.

CARVALHO, D.R.; MOSER, A.D.; SILVA, V.A.; DALLAGASSA, M.R. *Mineração de dados aplicada à fisioterapia. Fisioter Mov*, Curitiba, v.25, n.3, p.595-605, 2012.

FAYYAD, U.; SHAPIRO, G.P.S.; SMYTH, P. *From data mining to knowledge discovery in databases. American Association for Artificial Intelligence. 1996. Disponível em [http://courses.cse.tamu.edu/caverlee/csce670\\_2012/fayyad.pdf](http://courses.cse.tamu.edu/caverlee/csce670_2012/fayyad.pdf). Acesso em 11/11/2012.*

GUIMARÃES, L.G.C.; MORAES, J.V.; CARVALHO, D.R. *Data mining para avaliação do conhecimento sobre efeitos patrimoniais nas formas de união conjugal. In: IV Congresso de Tecnologias para gestão de dados e metadados do Cone Sul, Ponta Grossa, 2006. Disponível em <http://conged.deinfo.uepg.br/artigo3.pdf>. Acesso em 11/11/2012.*

HAN, J.; KAMBER, M. *Data Mining: concepts and techniques. Simon Fraser University. Morgan Kaufmann, 2000. Disponível em [http://www.cse.hcmut.edu.vn/~chauvtn/data\\_mining/Texts/\[1\]%20Data%20Mining%20-%20Concepts%20and%20Techniques%20\(3rd%20Ed\).pdf](http://www.cse.hcmut.edu.vn/~chauvtn/data_mining/Texts/[1]%20Data%20Mining%20-%20Concepts%20and%20Techniques%20(3rd%20Ed).pdf). Acesso em 11/11/2012.*

HUSSAIN, F.; LIU, H.; SUZUKI, E.; LU, H. *Exception rule mining with a relative interestingness measure. Lecture notes in Artificial Intelligence, 2000. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.34.1311>. Acesso em 12/11/2012.*

K/DOQI. KIDNEY DISEASE OUTCOMES QUALITY INITIATIVE. *Clinical Practice Guidelines for Cardiovascular Disease in Dialysis Patients. American Journal of Kidney Diseases*, v. 45, n. 4, p. S16, Apr, 2005.

K/DOQI. KIDNEY DISEASE OUTCOMES QUALITY INITIATIVE. *Clinical practice guidelines for chronic kidney disease: evaluation, classification and stratification. American Journal of Kidney Diseases*, v. 39, n. 2, p.1-246, 2002.

KIRSZTAJN, G.M.; BASTOS, M. *Proposta de uniformização da coleta de dados do previna-se. Oportunidade de conhecermos a parte submersa do “iceberg” Doença Renal Crônica no Brasil. Proposta de uniformização da coleta de dados do previna-se. J Bras Nefrol*, v. 29, n.4, dez, 2007.

KOTCHEN, T.A. *Doença vascular hipertensiva. In: Harrison medicina interna. 17a ed., Rio de Janeiro: Mc Graw Hill e Artmed, v.2, p.1549-1562, v.2, 2008.*

LEVIN, A.; HEMMELGARN, B.; CULLETON, B. et al. *Guidelines for the management of chronic kidney disease. Review. Canadian Medical Association Journal (CMAJ), 2008. Disponível em [www.cmaj.ca/content/179/11/1154.full.pdf](http://www.cmaj.ca/content/179/11/1154.full.pdf). Acesso em 12/08/2012.*

MILANI, C. S., CARVALHO, D. R. *Pós-Processamento em KDD. Revista de Engenharia e Tecnologia*. v.5, p.151 - 162, 2013.

MUNDSTOCK, E.; FACHEL, J.M.G.; CAMEY, A.S. et al. *Introdução à análise estatística utilizando o spss 13.0. Universidade Federal do Rio Grande do Sul. Instituto de Matemática. Cadernos de matemática e estatística. Série B: Trabalho de Apoio Didático. Porto Alegre, 2006. Disponível em [http://euler.mat.ufrgs.br/~giacomo/Manuais-oftw/SPSS/manual\\_spss\\_jandy.pdf](http://euler.mat.ufrgs.br/~giacomo/Manuais-oftw/SPSS/manual_spss_jandy.pdf). Acesso em 20/01/2013.*

NOGUEIRA, P.C.K.; FELTRAN, L.S.; CAMARGO, M.F. et al. *Prevalência estimada da doença renal crônica terminal em crianças no estado de São Paulo. Rev Assoc Med Bras, v.57, n.4, p. 443-449, 2011.*

ROMÃO JR, J.E. *Doença renal crônica: Definição, epidemiologia e classificação. Diretrizes brasileiras de doença renal crônica. J Bras Nefrol, v.26, n.1, p. 1-49, 2004.*

SAVIO, K.E.O.; COSTA, T.H.M.; SCHMITZ, B.A.S. et al. *Sexo, renda e escolaridade associados ao nível de atividade física de trabalhadores. Rev Saúde Pública, v.42, n.3, p. 457-463, 2008.*

SOCIEDADE BRASILEIRA DE CARDIOLOGIA. *VI diretrizes brasileiras de hipertensão. Arq Bras Cardio, v.95, n.1, p. 1-51, 2010.*

SOCIEDADE BRASILEIRA DE DIABETES. *Diretrizes da SBD/ 2012-2013. Métodos e critérios para o diagnóstico do diabetes mellitus. Capítulo 3. Disponível em <http://www.diabetes.org.br>. Acesso em 15/02/2013.*

SOCIEDADE BRASILEIRA DE DIABETES. *Posicionamento da SBD: algoritmo para tratamento da Diabetes Mellitus 2. Atualização, 2011. Disponível em <http://www.diabetes.org.br/>. Acesso em 12/11/2012.*

SOCIEDADE BRASILEIRA DE ENDOCRINOLOGIA E METABOLOGIA. *Projeto Diretrizes: sobrepeso e obesidade, 2004. Disponível em <http://www.endocrino.org.br/>. Acesso em 11/11/2012.*

SOCIEDADE BRASILEIRA DE NEFROLOGIA. *Campanhas. Previna-se. 2010. Disponível em <http://www.sbn.org.br/index.php?previna&menu=6>. Acesso em 12/08/2012.*

SOCIEDADE BRASILEIRA DE NEFROLOGIA. *Censo 2013. Disponível em <http://www.sbn.org.br/leigos/index.php?censo>. Acesso em 17/06/2014.*

STEPHEN, R.G. *Weka: The Waikato Environment for Knowledge Analysis. 2010. Disponível em <http://www.cs.waikato.ac.nz/~ml/publications/1995/Garner95-WEKA.pdf>. Acesso em 15/01/2013.*

TOSCANO, C.M. *As campanhas nacionais para detecção das doenças crônicas não-transmissíveis: diabetes e hipertensão arterial. Ciências & Saúde Coletiva, v.9, n.4, p.885-895, 2004.*