

ALGORITMOS DE MINERAÇÃO DE DADOS EM SISTEMA DE MONITORAMENTO DE DIABETES

**Robson Cezário¹; Eduardo Palharini²;
Franceliane Jobim Benedetti³; Gedson Leal Gomes⁴; Sylvio Vieira⁵;
Robertson E. dos Santos⁶; Alexandre Zamberlan⁷**

RESUMO

Este trabalho está no contexto de extração e descoberta de conhecimento em bases de dados referentes às pessoas com diabetes. A base contém dados de monitoramento de data, hora e quantidade de insulina utilizada, quantidade de calorias e carboidratos ingeridos naquela data, qualidade do sono (escala de Likert) e tempo em minutos de atividades físicas (escala de Borg). Contudo, uma base com muitos dados, há dificuldade de se encontrar padrões de forma visual, mesmo que por relatórios. Portanto, esta pesquisa busca estudar, aplicar, avaliar e apontar métodos de mineração de dados adequados para esse contexto. A pesquisa é baseada em revisão bibliográfica e estudo de caso, em que o trabalho de (RATH; ZAMBERLAN; VIEIRA, 2014) é estendido e integrado com uma base de três (3) anos de dados coletados em tabela de planilha eletrônica. E por meio dessa nova base, os algoritmos de mineração devem ser avaliados.

Palavras-chave: Descoberta de conhecimento; Sistema de comportamento inteligente; .

Eixo Temático: Tecnologia, Inovação e Desenvolvimento Sustentável.

1. INTRODUÇÃO

A área da Saúde produz muitos e variados dados, sejam eles na Farmácia, Medicina, Fisioterapia, Enfermagem, entre outros. Tratamento de pacientes com diabetes, por exemplo, também acabam gerando uma quantia significativa e diversificada de dados, principalmente, quando pacientes, médicos, farmacêuticos e nutricionistas monitoram a evolução (ou controle) da doença. Esse controle pode ser

¹ Apresentador - Sistemas de Informação UFN - robson.cezario@ufn.edu.br

² Bolsista Sistemas de Informação UFN - epavani@ufn.edu.br

³ Professora Nutrição UFN - franceliane@ufn.edu.br

⁴ Professor de Educação Física e Personal Trainer - gedson.gomes@gmail.com

⁵ Professor Coordenador do Laboratório de Práticas da Computação - sylvio@ufn.edu.br

⁶ Analista de Sistemas da ER Sistemas - robertson@ersistema.info

⁷ Professor Ciência da Computação e Laboratório de Práticas UFN - alexz@ufn.edu.br

via monitoramento de valores diários de unidades de insulina aplicada, valores diários de glicemia em jejum, quantidade de calorias e carboidratos consumidos, tempo e quantidade de exercícios realizados, entre outros. Entretanto, muitos e diferentes dados dificultam a análise e o reconhecimento de padrões que possam estar embutidos (até mesmo ocultos) numa base de dados. Assim, médicos, nutricionistas e até educadores físicos podem elaborar estratégias falhas, ou incompletas, pois a análise acaba sendo superficial.

Em 2014, um aluno do curso de Sistemas de Informação da UFN (RATH; ZAMBERLAN; VIEIRA, 2014) projetou e implementou um Sistema de Recomendação para Diabetes, construído via a linguagem PHP e banco de dados MySQL. O projeto disponibilizou um sistema web para que pacientes diabéticos, seus médicos e nutricionistas pudessem registrar dados de insulina, glicemia, calorias, carboidratos, exercícios, qualidade de sono, etc. Porém, o projeto aplicou parcialmente técnicas de mineração para descoberta de padrões. O projeto trabalhou com mineração, mas não era o foco principal, ou seja, o foco era o sistema web para registro e acompanhamento de dados de pessoas diabéticas. A mineração foi prototipada mais em um sentido de criar algo funcional com o mínimo possível para ser analisado e testado. Dessa forma, este trabalho pretende estender a pesquisa realizada em (RATH; ZAMBERLAN; VIEIRA, 2014), principalmente na análise e aplicação de melhores técnicas de mineração de dados na base projetada para diabetes. Registra-se que há uma base de dados criada e populada entre os anos 2012 e 2014 (independente do sistema projetado em (RATH; ZAMBERLAN; VIEIRA, 2014)), com 722 dias monitorados. E essa base será utilizada para o estudo, aplicação e avaliação das diferentes técnicas de mineração.

O objetivo inicial é estudar, comparar, aplicar e avaliar algoritmos presentes no ambiente WEKA para reconhecer precisamente padrões presentes na base de dados do estudo. Objetivos específicos são: i) entender e aplicar as categorias de algoritmos de mineração de dados aos diferentes contextos (TAN; STEINBACH; KUMAR, 2009), (MARQUES et al., 2008); ii) estudar e avaliar algoritmos presentes

no ambiente WEKA (GROUP, 2021); iii) realizar adequação da base de dados do sistema projetado por (RATH; ZAMBERLAN; VIEIRA, 2014); iv) mapear e compilar trabalhos relacionados que usaram técnicas de mineração em bases de dados (MARQUES et al., 2008), (VIEIRA, 2016); v) definir, aplicar, analisar um estudo de caso para validar a proposta.

2. REVISÃO BIBLIOGRÁFICA

As áreas de conhecimento produzem, diariamente, uma quantidade de dados e de informação muito grande. Esses dados e informações precisam ser coletados, tratados e armazenados, produzindo assim uma base de dados e/ou conhecimento específico.

Por exemplo, no campo da Medicina, dados coletados não são o suficiente para uma tomada de decisão precisa e eficiente. Assim, para se ter uma decisão mais assertiva, são necessárias ferramentas que facilitam e auxiliam a verificação/análise desses dados. Logo, conforme apresentado em (RATH; ZAMBERLAN; VIEIRA, 2014), boas tomadas de decisão em relação aos quadros clínicos de pacientes diabéticos, por exemplo, influenciam diretamente no tratamento e na prevenção da doença. A análise de dados nas áreas da Saúde tem sido realizada por meio estatístico, que usa um processo matemático estabelecido com suporte teórico que permite interpretações. No entanto, há um método alternativo que auxilia em análises e interpretações de grandes quantidades de dados, conhecido como Mineração de Dados (*Data Mining* - DM). Esse método, por meio do uso ou não da estatística, busca encontrar modelos ou padrões ocultos dentro de uma base, que dificilmente seriam detectados com técnicas estatísticas. De acordo com o trabalho de (RATH; ZAMBERLAN; VIEIRA, 2014), alguns autores consideram os termos Mineração de Dados e Descoberta de Conhecimento em banco de dados (*Knowledge-Discovery in Databases* - KDD) como processos diferentes. De fato, mineração faz parte da linha de descoberta de conhecimento. Entretanto, neste trabalho, assume-se que os termos têm o mesmo significado: extrair conhecimento

de dados.

A definição aceita e citada por vários autores sobre mineração, segundo (REZENDE, 2003), é: “extração de conhecimento de base de dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados”. No trabalho de (RATH; ZAMBERLAN; VIEIRA, 2014), a mineração foi dividida em 3 etapas: coleta de dados, processamento e identificação de padrões, e pós-processamento.

2.1 Algoritmos clássicos para mineração de dados

De acordo com (LAROSE, 2005), mineração de dados é classificada conforme as tarefas realizadas, sendo as mais comuns: Descrição: tarefa que descreve os padrões descobertos e disponibiliza uma possível interpretação dos resultados, via técnicas exploratórias de dados; Classificação: tem como objetivo detectar a qual classe uma determinada informação pertence, como um processo de categorização, geralmente realizada por processo de aprendizado supervisionado; Regressão: semelhante à classificação, mas é utilizada quando a informação é identificada por um valor numérico e não categórico; Predição: também semelhante à classificação e regressão, no entanto busca descobrir o valor futuro de um determinado atributo; Agrupamento: identifica e aproxima informações semelhantes. Um agrupamento é um conjunto de informações similares entre si, mas diferentes de outras informações nos demais agrupamentos; Associação: identifica a relação entre atributos de um conjunto de informações. Segundo o trabalho realizado em (FURLAN; SOUZA POLETTTO, 2018), a mineração de dados possui várias implementações distintas por meio de diversos algoritmos. Esses algoritmos são segmentados, novamente, pelas tarefas. A Figura 1 mostra uma relação dos principais algoritmos, suas descrições, tarefas e exemplos.

Técnica	Descrição	Tarefas	Exemplos
Árvore de Decisão	Baseada em estágios de decisão (nós) e na separação de classes e subconjuntos, organiza os dados de forma hierárquica.	- Classificação - Predição	CART, CHAID, C5.0, ID-3
Redes Neurais	Modelos inspirados na fisiologia do cérebro, nos quais o conhecimento é fruto do mapa de conexões neuronais e dos pesos dessas conexões.	- Classificação - Agrupamento - Predição	Perceptron, Rede MLP, Redes de Kohonen, Rede Hopfield, Rede BAM, Redes ART, Rede IAC, Rede LVQ, Rede Counterpropagation, Rede RBF, Rede PNN, Rede Time Delay, Neocognitron, Rede BSB.
Raciocínio Baseado em Casos	Baseado no método do vizinho mais próximo combina e compara atributos para estabelecer hierarquia de semelhança.	- Classificação - Agrupamento	BIRCH, CLARANS CLIQUE
Algoritmos Genéticos	Métodos gerais de busca e otimização, inspirados na Teoria da Evolução, em que a cada nova geração, soluções melhores têm mais chance de ter "descendentes".	- Classificação - Agrupamento	Algoritmo Genético Simples, Genitor, GA-Nuggets, GAPVMINER
Conjuntos Fuzzy	Oferece uma grande vantagem para classificar dados com um alto nível de abstração.	- Classificação - Agrupamento	K-means, FCMdd
Regras de Indução	Processo para obter uma hipótese a partir de dados e fatos já existentes.	- Classificação - Predição	CART, CHAID
Regras de Associação	Estabelece uma correlação estatística entre atributos de dados e conjuntos de dados.	- Associação	Apriori, AprioriTid, AprioriHybrid, AIS, SETM

Figura 1 - Técnicas e Tarefas empregadas na Mineração de Dados (GOLDSCHMIDT, 2005).

2.2 Ambientes ou ferramentas para Mineração de Dados

Um dos ambientes mais conhecidos e utilizados em instituições de ensino e de pesquisa, no contexto de descoberta de conhecimento e mineração de dados, é o *Waikato Environment for Knowledge Analysis* (WEKA). É uma ferramenta de código aberto, com licença *General Public Licence* (GPL), multiplataforma (Windows, MAC OS e Linux), contendo inúmeras funções, como por exemplo, a importação de banco de dados em arquivo para mineração (.arff). Além disso, possui diversos algoritmos de mineração que podem ser aplicados, testados e analisados em diferentes bases importadas. O WEKA, enquanto ferramenta, possui ambientes de importação, de visualização de resultados e de algoritmos de mineração, como algoritmos de redes neurais para aprendizado de máquina (*machine learning*), entre outros. No ambiente de algoritmos para mineração, há sub-ferramentas de pré-processamento, oferecendo suporte para todo processo de mineração, incluindo a preparação dos dados de entrada, avaliação estatística de aprendizagem, visualização dos dados de entrada e seus resultados. A ferramenta WEKA foi desenvolvida na linguagem JAVA, em que é possível tanto usar uma interface gráfica específica dela, quanto importar pacotes WEKA, contendo os algoritmos de mineração, em aplicações JAVA.

WEKA possui funcionalidades divididas em ambientes, que são chamadas em

botões na interface principal, como: *Explorer*: possui aplicações de pré-processamento, análise e visualização dos resultados; *Experimenter*: aqui, o usuário pode realizar testes estatísticos entre as estruturas de aprendizagem da ferramenta, onde se pode utilizar vários algoritmos concomitantes e comparar os resultados, escolhendo assim o algoritmo mais adequado para a base; *Knowledge Flow*: com funções similares ao *Explorer*, diferencia-se pela representação gráfica dos resultados; *Workbench*: ambiente que combina todas as interfaces gráficas em uma única interface; *Simple CLI*: ambiente para inserção de comandos em forma de *shell* ou terminal. Por exemplo, após o banco de dados ser carregado é possível realizar a edição para uma eventual correção e filtrar pelos atributos conforme ilustrado na Figura 2.

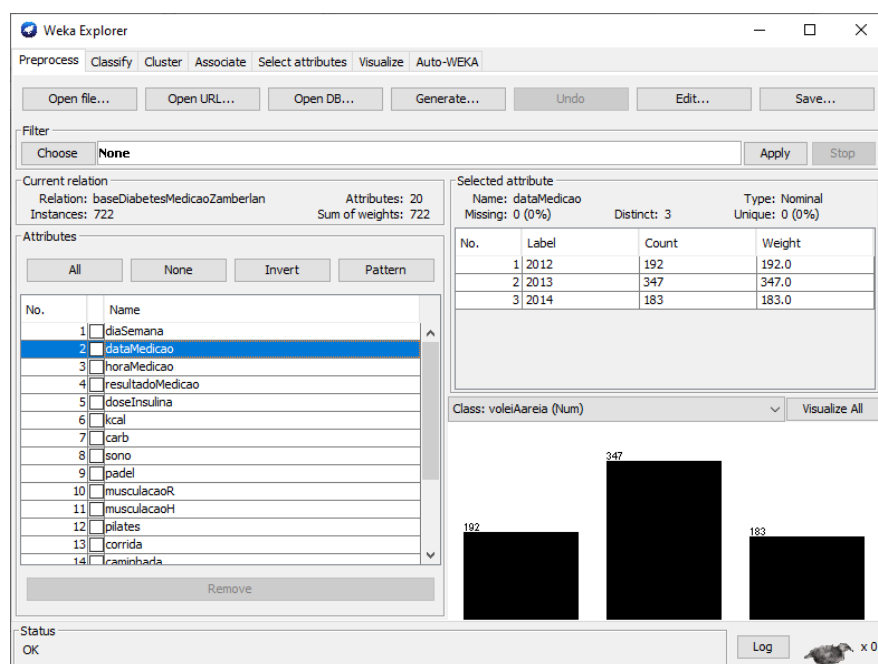


Figura 2 - WEKA Explorer com a base de dados carregada.

2.3 Trabalhos relacionados

O primeiro trabalho é a base para o estudo de caso desta pesquisa, onde Rath e colaboradores (2014) desenvolveram um sistema Web, na linguagem PHP, para cadastro e controle diário de dados do diabético, que são armazenados em banco de dados MySQL. O trabalho utilizou algoritmos presentes no ambiente WEKA de mineração.

No trabalho realizado em (MARQUES et al., 2008), foram aplicadas técnicas de mineração de dados para um sistema de apoio à tomada de decisão no contexto de monitoramento de jogadas em partidas de Voleibol. O sistema é conhecido como *scout*, com apelo estatístico, que possui informações muitas vezes desnecessárias, que na mineração de dados são descartadas. Dessa forma, o processo de mineração no *scout* de Voleibol foi útil para destacar informações confiáveis e algumas até desconhecidas, assim ajudando de maneira mais eficaz as comissões técnicas. O sistema foi construído na linguagem de programação Java e com o apoio do ambiente WEKA.

Segundo o que foi apresentado em (FURLAN; SOUZA POLETTTO, 2018), um estudo foi realizado para aplicar conceitos de descoberta de conhecimento e de mineração em banco de dados. Alguns algoritmos do ambiente WEKA para visualização do processo foram testados.

Analisando os trabalhos, em (RATH; ZAMBERLAN; VIEIRA, 2014), havia o objetivo de aplicar técnicas de mineração de dados para encontrar padrões na sua base, para poder recomendar alimentos e atividades físicas aos diabéticos. No trabalho, foi projetado e implementado um sistema Web para isso. Porém, essa recomendação não foi totalmente finalizada, uma vez que a base de dados não tinha um conjunto de dados significativo para que algoritmos de mineração fossem utilizados. Dessa forma, a principal justificativa deste trabalho. No trabalho de (MARQUES et al., 2008), destaca-se o processo de limpeza da base, a revisão bibliográfica das categorias dos diferentes algoritmos de mineração, facilitando a escolha de algoritmos para esta pesquisa. Já no trabalho de Furlan e Souza Poletto (2018), foi mostrada a análise de técnicas de algoritmos de mineração de dados presente no ambiente WEKA. Também compilou informações sobre mineração de dados e descoberta de conhecimento. Dessa forma, ficou evidente o uso da ferramenta WEKA de forma simplificada.

2.4 Monitoramento da diabetes

De acordo com a Sociedade Brasileira de Diabetes (SBD, 2021), a diabetes

mellitus é uma doença crônica em que o corpo não é capaz de produzir ou absorver corretamente a insulina gerada pelo pâncreas. A insulina é o hormônio responsável pelo controle da quantidade de glicose no sangue que um ser humano recebe na alimentação como fonte de energia para o organismo (SBD, 2021). Novamente conforme a SBD, quando o ser humano tem diabetes e caso não for controlada de forma adequada por longos períodos, podem aparecer complicações, como doenças renais, infarto do miocárdio, acidente vascular cerebral, pé diabético, glaucoma, catarata, entre outros (SBD, 2021). De acordo com o portal da secretaria de atenção primária à saúde, estima-se que no Brasil 9 milhões de pessoas com diabetes (<https://aps.saude.gov.br/noticia/10336>). Cabe ressaltar que o tratamento com insulina, busca, de forma artificial e externa ao corpo, a regulação desse hormônio, para que cumpra seu papel no organismo. Sendo assim, todos portadores de diabetes Tipo 1 e Tipo 2 podem necessitar de insulina para controlar a glicose no sangue. Para essa finalidade, existem vários tipos de insulina disponíveis para o tratamento. E tal diferença se dá pelo tempo que ficam ativas no corpo, pelo tempo que levam para agir e em qual situação do dia são mais eficientes.

Independente do tipo de diabetes ou do tipo de terapia ou tratamento, o monitoramento é essencial para o controle, para o tratamento, para o acompanhamento dos profissionais Médicos, Nutricionistas e Educadores Físicos. Há profissionais da área da Saúde que monitoram seus pacientes com exames de sangue tradicionais, realizados de tempos em tempos, como glicemia em jejum, glicemia glicada, insulina, triglicerídeos, colesterol, etc. Porém, há outros tipos de acompanhamento, ditos diários: glicemia, unidades aplicadas de insulina, qualidade do sono, quantidade de calorias e carboidratos ingeridos, tempo e tipo de atividade física.

2. METODOLOGIA

Este trabalho é baseado em pesquisa exploratória com revisão bibliográfica amparado com estudo de caso. O estudo de caso é aplicação de mineração na base

de dados criada entre 2012 e 2014. Já no projeto e desenvolvimento da solução, são utilizados a metodologia *Scrum* com a técnica *Kanban*. As ferramentas utilizadas foram: Trello - técnica kanban; GitHub - controle de versão; Astah - diagramação UML; Ambiente WEKA - simulador para testes de algoritmos de mineração de dados.

A Figura 3 mostra como os dados monitorados (dias da semana, ano, glicemia, quantidade de insulina, quantidade calorias, quantidade de carboidratos, qualidade do sono, tempo de atividades físicas) foram coletados. O destaque é que a média de glicemia era analisada sempre de 3 em 3 dias, pois essa média sugeria alteração nas unidades de insulina aplicada. Ou seja, se a média fosse maior que 100 mg/dL, a quantidade de unidades de insulina deveria ser aumentada em duas. Caso fosse menor que 80 mg/dL deveria diminuir em 2 unidades.

	A	B	C	D	E	F	G	H	I	J	K
483	Sábado	4/13/2013	7:00	ac	99	6	2878	317	2	80 min padel	
484	Domingo	4/14/2013	12:00	ac	101	6	2280	252	4		
485				média	100		2,305	251			
486	Segunda	4/15/2013	7:00	ac	99	6	2835	287	3	60 min padel	
487	Terça	4/16/2013	7:00	ac	102	6	2404	260	4	30 min musculação R	
488	Quarta	4/17/2013	7:00	ac	111	8	2181	237	3	90 min padel	
489				média	104		2,473	261			
490	Quinta	4/18/2013	7:20	ac	88	8	2141	240	5		
491	Sexta	4/19/2013	11:00	ac	99	8	2405	241	3		
492	Sábado	4/20/2013	9:30	ac	99	8	2715	315	3	40 min padel	
493				média	95		2,420	265			
494	Domingo	4/21/2013	10:00	ac	99	8	2407	237	4	150 min padel	
495	Segunda	4/22/2013	6:55	ac	94	8	1936	220	3		
496	Terça	4/23/2013	7:20	ac	90	8	1697	237	4		
497				média	94		2,013	231			
498	Quarta	4/24/2013	7:00	ac	94	8	2096	282	4		
499	Quinta	4/25/2013	7:00	ac	101	8	2419	215	4		
500					98	8	2360	204	3	50 min pilates 20 min bike	troquei agulha e lanceta
501	Sexta	4/26/2013	7:00		98		2,292	227			
502	Sábado	4/27/2013	7:00	ac	94	8	1945	153	4	20 min caminhada 30 min pilates 20 caminhada	
503	Domingo	4/28/2013	9:45	ac	101	8	2296	239	4	60 min padel	
504	Segunda	4/29/2013	6:45	ac	92	8	1937	182	5		
505				média	96		2,059	191			

Figura 3 - Versão inicial dos dados armazenados em planilha eletrônica.

Destaca-se, também, que a plataforma WEKA precisa receber uma base tratada, preferencialmente com valores categóricos, assim foi preciso melhorar alguns campos (colunas ou atributos). Dessa forma, foram realizadas conversões de valores numéricos para valores categóricos, principalmente nos campos quantidade de exercícios realizados, quantidade de carboidratos, quantidade de calorias ingeridos em um dia e resultado da medição da glicemia, como mostra a Figura 4.

Em relação à quantidade de exercícios, optou-se em utilizar a escala (ou tabela) de Borg (de 0 a 20), que trabalha com uma percepção subjetiva de quanto esforço o próprio atleta realizou, sendo que 1 é quase nenhum esforço e 20 o

esforço máximo possível (BORG; BORG, 2001). De acordo com Borg (2001), os valores entre 6 e 20 são baseados na Frequência Cardíaca de 60 a 200 batimentos por minuto (bpm), sendo que o valor 12 corresponde aproximadamente 55% e o 16 a 85% da Frequência Cardíaca Máxima. Dessa forma, a escala corresponderia ao esforço da seguinte forma: 7 - muito fácil; 9 - fácil; 11 - relativamente fácil; 13 - ligeiramente cansativo; 15 - cansativo; 17 - muito cansativo; 19 - exaustivo.

Em relação ao consumo de carboidratos e calorias, também foram necessárias modificações na tabela original, transformando dados numéricos em dados categóricos, como "abaixo", "recomendado" e "acima" do recomendado. Dessa forma, se uma pessoa tem como orientação (referência), na sua dieta, ingerir no máximo 200 gramas carboidratos e 2600 calorias, o abaixo do recomendado é quando essa pessoa ingere menos de 80% desses valores de referência, recomendado quando ingere entre 80% a 120% dos valores e acima do recomendado quando consome mais de 120% dos valores sugeridos na dieta. Essas orientações são estratégias básicas no contexto da nutrição. Por fim, em relação à resultados de medição de glicemia há as categorias "abaixo", "recomendado" e "acima" do recomendado. Porém, há dois momentos de medição: em jejum e pós-prandial (após 2 horas de uma refeição). Dessa forma, para medições em jejum, o abaixo do recomendado é um valor menor que 80 mg/dL (miligramas por decilitro), recomendado é um valor entre 80 mg/dL e 100 mg/dL e acima do recomendado um valor acima de 100 mg/dL. Para medições pós-prandial, o recomendado é um valor abaixo de 140 mg/dL e acima, um valor maior que 140 mg/dL.

Dia Semana	Data	Antes Comer / Depois Comer	Resultado	Dose Insulina	kcal	carb	noite de sono 1-5	padel	musculação R	musculação H	pilates	corrida	caminhada	tênis	sauna	bike	natação	elíptico	volei de areia
Quinta	2013	ac	99	6	Recomendado	Recomendado	3	11	0	0	0	0	0	0	0	0	0	0	0
Sexta	2013	ac	97	6	Recomendado	Acima	3	0	0	0	0	0	0	0	0	0	9	0	0
Sábado	2013	ac	98	6	Recomendado	Acima	3	0	0	0	0	0	0	13	0	11	0	0	0
Domingo	2013	ac	98	6	Recomendado	Acima	3	0	0	0	0	0	0	15	0	0	0	0	0
Segunda	2013	ac	92	6	Recomendado	Acima	4	15	0	0	0	0	0	0	0	0	0	0	0
Terça	2013	ac	106	6	Abaixo	Recomendado	3	0	0	0	0	0	0	0	0	0	0	0	0
Quarta	2013	ac	100	6	Recomendado	Acima	4	17	11	0	0	0	0	0	0	0	0	0	0
Quinta	2013	ac	91	6	Recomendado	Recomendado	4	0	0	0	0	0	0	0	0	0	0	0	0
Sexta	2013	ac	89	6	Recomendado	Acima	3	11	0	0	13	0	0	0	0	0	0	0	0
Sábado	2013	ac	100	6	Recomendado	Recomendado	4	0	0	0	0	0	0	15	0	0	0	0	0
Domingo	2013	ac	92	6	Recomendado	Acima	4	0	0	0	0	0	0	15	0	0	0	0	0
Segunda	2013	ac	95	6	Recomendado	Recomendado	4	15	0	0	0	0	0	0	0	0	0	0	0

Figura 4 - Versão adaptada da tabela para ser importada no WEKA.

Uma vez melhorada a tabela, já é possível no ambiente WEKA importar a base da planilha eletrônica. A importação (automática) converte a tabela em um arquivo com extensão *.arff*, com os campos (colunas) para sua manipulação interna, conforme mostra a Figura 5.

```

1 @attribute dia {Quinta,Sexta,Sabado,Domingo,Segunda,Terca,Quarta}
2 @attribute ano numeric
3 @attribute 'hora comida' {ac}
4 @attribute glicemia numeric
5 @attribute 'insulina' numeric
6 @attribute kcal {Abaixo,Recomendado,Acima}
7 @attribute 'carb' {Acima,Recomendado,Abaixo}
8 @attribute sono numeric
9 @attribute padel numeric
10 @attribute musculacaoR numeric
11 @attribute musculacaoH numeric
12 @attribute pilates numeric
13 @attribute corrida numeric
14 @attribute caminhada numeric
15 @attribute tenis numeric
16 @attribute sauna numeric
17 @attribute bike numeric
18 @attribute natacao numeric
19 @attribute eliptico numeric
20 @attribute voleiareaia numeric
21 @data
22 Quinta,2012,ac,90,6,Abaixo,Acima,4,0,0,0,0,0,0,0,0,0,0,0,0,0
23 Sexta,2012,ac,96,6,Recomendado,Acima,4,0,13,0,0,11,0,0,0,0,0,0,0
24 Sabado,2012,ac,90,6,Recomendado,Acima,4,17,0,0,0,0,0,0,0,0,0,0,0
25 Domingo,2012,ac,105,6,Abaixo,Acima,5,0,0,0,0,0,0,0,0,0,0,0,0
26 Segunda,2012,ac,86,6,Recomendado,Acima,5,17,0,0,0,0,0,0,0,0,0,0,0
27 Terca,2012,ac,86,6,Recomendado,Acima,4,0,0,0,0,0,0,0,0,0,0,0,0
28 Quarta,2012,ac,97,6,Recomendado,Acima,5,0,0,0,0,0,0,13,0,0,0,0,0

```

Figura 5 - Arquivo arff da conversão da planilha eletrônica para a ferramenta WEKA.

Ressalta-se que *@attribute* representa cada coluna da planilha eletrônica. Alguns atributos precisam ser valorados, como por exemplo, o atributo *dia* que contém os dias da semana. Finalmente, a partir da linha 22, há os dados da planilha convertidos para a plataforma WEKA. Por exemplo, na linha 25 há uma medição referente a um domingo, do ano de 2012, com uma medição antes de comer, com 100 mg/dL de glicemia, 6 unidades de insulina aplicada, com ingestão abaixo do recomendado para calorias, acima do recomendado para ingestão de carboidratos, 5 pontos na qualidade do sono e nenhuma atividade física realizada.

3. CONCLUSÃO

Esta pesquisa, encontra-se em fase inicial, mas com toda a preparação da base (adaptada) para que os algoritmos de mineração sejam aplicados via WEKA. Uma vez que isso aconteça, será possível identificar padrões como: se a atividade

foi mais aeróbica, tipo corrida, jogo de tênis, como se deu a dieta e como isso impacta a quantidade de glicemia e a qualidade do sono.

REFERÊNCIAS

- BORG, G.; BORG, E. A new generation of scaling methods: Level-anchored ratio scaling. *Psychologica*, 28:15-45, 2001.
- FURLAN, M. B.; SOUZA POLETTTO, A. S. R. Algoritmos e técnicas para mineração de dados. Fundação Educacional do Município de Assis/SP, 2018.
- GOLDSCHMIDT, E. P. R. Data Mining: Um Guia Prático. Editora Elsevier, 2005.
- GROUP, M. L. WEKA: The workbench for machine learning. <https://www.cs.waikato.ac.nz/ml/weka/>. Acessado em Março de 2021.
- LAROSE, D. T. Discovering Knowledge in Data: An Introduction to Data Mining. Editora John Wiley and Sons, 2005.
- MARQUES, É. B.; ZAMBERLAN, A.; OLIVEIRA, R. F.; Raimann, L. H.; OLIVEIRA, V. Projeto de módulo de data mining para scout voleibol. Seminário de Informática do Rio Grande do Sul (SEMINFO RS 2008), Torres, RS, 2008.
- RATH, R.; ZAMBERLAN, A.; VIEIRA, S. Sistema de recomendação para controle da diabetes. 7º Congresso Sul Brasileiro de Computação, Criciúma. SULCOMP, UNESC, 2014.
- REZENDE, S. O. Sistemas inteligentes: fundamentos e aplicações. Editora Manole Ltda, 2003.
- SBD. Sociedade Brasileira de Diabetes. <https://www.diabetes.org.br/>. Acessado em Março de 2021.
- TAN, P. N.; STEINBACH, M.; KUMAR, V. Introdução ao datamining: mineração de dados. Ciência Moderna, São Paulo, 2009.
- VIEIRA, S. A. G. Identificação de padrões de expressão em doenças genéticas usando uma rede de integração de vias de manutenção do Genoma, Angiogênese, Hipóxia e Vigilância Imunológica. Tese de Doutorado, UNIFRA, Santa Maria, 2016.