

UnB - UNIVERSIDADE DE BRASÍLIA
FGA - FACULDADE GAMA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

**PROGRAMA DE EDUCAÇÃO EM DIABETES COM ÊNFASE EM
EXERCÍCIOS FÍSICOS: ANÁLISE E MINERAÇÃO DE DADOS**

THIANE MARQUES TORQUATO

ORIENTADORA: Dra. Suélia de Siqueira Rodrigues Fleury Rosa

CO-ORIENTADORA: Dra. Juliana de Faria Fracon e Romão

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA BIOMÉDICA

PUBLICAÇÃO: 45A/2016

BRASÍLIA/DF: MAIO– 2016

UnB - UNIVERSIDADE DE BRASÍLIA

FGA - FACULDADE GAMA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA

**PROGRAMA DE EDUCAÇÃO EM DIABETES COM ÊNFASE EM EXERCÍCIOS
FÍSICOS: ANÁLISE E MINERAÇÃO DE DADOS**

THIANE MARQUES TORQUATO

DISSERTAÇÃO DE MESTRADO ACADÊMICO SUBMETIDA AO PROGRAMA DE ENGENHARIA BIOMÉDICA DA FGA - FACULDADE GAMA, DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA BIOMÉDICA.

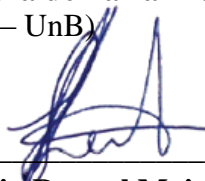
APROVADA POR:



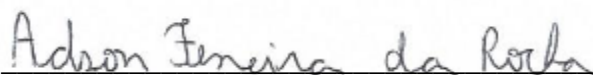
Prof^a Dra. Suévia de Siqueira Rodrigues Fleury Rosa
(Orientadora - UnB)



Prof^a Dra. Juliana de Faria Fracon e Romão
(Co-orientadora – UnB)



Prof^a Dra. Cícilia Raquel Maia Leite
(Examinador Externo - UERN)



Prof Dr. Adson Ferreira da Rocha
(Examinador Interno - UnB)

BRASÍLIA/DF, 16 DE ABRIL de 2016.

BRASÍLIA/DF, 16 DE ABRIL de 2016.

FICHA CATALOGRÁFICA

THIANE MARQUES TORQUATO

Programa de Educação em Diabetes com Ênfase em Exercícios Físicos: Análise e Mineração de Dados, [Distrito Federal] 2016.

97p., 210x297 mm (FGA/UnB Gama, Mestre, Engenharia Biomédica, 2016)

1. Diabetes Mellitus. 2. Mineração de Dados

3. Educação em Saúde. 4. Exercício Físico

I. FGA UnB Gama/ UnB. II. Programa de Educação em Diabetes com Ênfase em Exercícios: Análise e Mineração de Dados - Dissertação de Mestrado - Universidade de Brasília. Faculdade Gama - Programa de Pós Graduação em Engenharia Biomédica

REFERÊNCIA BIBLIOGRÁFICA

TORQUATO, T. M. (2016). PROGRAMA DE EDUCAÇÃO EM DIABETES COM ÊNFASE EM EXERCÍCIOS FÍSICOS: ANÁLISE E MINERAÇÃO DE DADOS. Dissertação de Mestrado em Engenharia Biomédica, Publicação N°. 45A/2016, Programa de Pós-graduação em Engenharia Biomédica, Faculdade Gama, Universidade de Brasília, Brasília, DF, 97 p.

CESSÃO DE DIREITOS

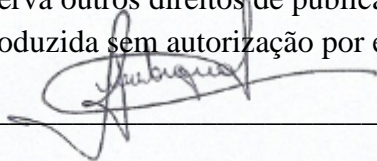
AUTOR: THIANE MARQUES TORQUATO

TÍTULO: PROGRAMA DE EDUCAÇÃO EM DIABETES COM ÊNFASE EM EXERCÍCIOS FÍSICOS: ANÁLISE E MINERAÇÃO DE DADOS

GRAU: MESTRE

ANO: 2016

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.



Thiane Marques Torquato

Rua Maringá Qd. 31 Lt. 12- Setor Casego. CEP: 76400-000- Uruaçu- GO - BRASIL

DEDICATÓRIA

Aos meus pais e irmãs;

Ao meu marido e filho;

E principalmente a Deus: razão
e origem de tudo o que somos e fazemos.

AGRADECIMENTOS

À Deus, que é vida.

À minha orientadora, Profa. Dra. Suélia Fleury, pela dedicação, pela amizade, obrigada pela paciência e pelo carinho que sempre demonstrou ao me orientar, por acreditar que é possível e tê-la conhecido foi uma honra, tê-la no meu convívio é um presente maravilhoso!

À minha co-orientadora, Profa. Dra. Juliana Fracon pelo apoio, atenção e pelas valiosas sugestões.

Ao Instituto Federal de Goiás – Campus Uruaçu pelo apoio nos momentos em que se fizeram necessários na condução do mestrado e a CAPES pela bolsa concedida.

Ao meu pai Djalma e minha mãe Luzeni, pelas orações, por fazer parte de tudo o que eu sou, por contribuir para que eu aprendesse a observar a vida e escolher meu caminho e por estar sempre presente, em todos os momentos, apoiando-me.

Ao meu amigo Gilderlan que nesses 2 anos de mestrado me acolheu em sua residência em Brasília. Tenho imensa gratidão por você.

Às minhas queridas irmãs Thiara, Thatiane e Diandra por fortalecerem minha alegria de viver.

Ao meu esposo Lynwood, por compreender os momentos de ausência dedicados aos estudos, pela colaboração e ajuda na educação do nosso filho, meu companheiro e incentivador, sempre compreensivo e, mesmo nos momentos mais difíceis, soube acolher minhas incertezas e com muito carinho soube me apoiar e tranquilizar. Obrigada por tudo!

À Direção do Programa Doce Desafio pelo incentivo à realização deste trabalho, e especialmente a Profa. Dra. Jane Dullius por sua contribuição profissional ao dedicar seu tempo para sanar minhas dúvidas relacionadas à área da saúde.

Aos amigos de mestrado Leandra e Gustavo, pelos risos, pelas discussões enriquecedoras, pela presença e pela amizade.

A todos os alunos do IFG que tive e terei, sem os quais o ato de ensinar/aprendendo e aprender/ensinando, não teria sentido.

À razão de todas as minhas buscas, luz da minha vida... meu filho Wood, fonte inesgotável de inspiração e coragem.

“Estamos afogados em informação, mas morrendo de fome por conhecimento.”

- John Naisbett

RESUMO

PROGRAMA DE EDUCAÇÃO EM DIABETES COM ÊNFASE EM EXERCÍCIOS FÍSICOS: ANÁLISE E MINERAÇÃO DE DADOS

Autor: Thiane Marques Torquato

Orientadora: Prof^a Dra. Suélia de Siqueira Rodrigues Fleury Rosa

Co-orientadora: Prof^a Dra. Juliana de Faria Fracon e Romão

Programa de Pós-Graduação em Engenharia Biomédica

Brasília, maio de 2016.

O *Diabetes Mellitus* (DM) é uma doença metabólica com causas multifatoriais relacionada diretamente a alterações no pâncreas devido à incapacidade de secretar insulina (*Diabetes Mellitus* tipo 1 - DM1) e/ou produzir insulina (*Diabetes Mellitus* tipo 2 - DM2). A principal maneira de diminuir significativamente as complicações do DM é a adesão as orientações dietéticas, de exercício, ao tratamento medicamentoso e ao automonitoramento glicêmico com apoio dos programas de educação em saúde. O Programa Doce Desafio (PDD) é um programa de educação em saúde da Universidade de Brasília (UnB) que se iniciou em 2001 e vem sendo realizado de forma contínua que visa alcançar a melhoria da qualidade de vida das pessoas diabéticas, por meio da prática orientada de exercícios físicos. Nas aulas de educação física é realizado o acompanhamento dos pacientes através de preenchimento de fichas diárias, em consequência grandes volumes de dados têm sido coletados e armazenados. Assim, para auxiliar na investigação desses materiais, objetiva-se neste estudo utilizar as técnicas da mineração de dados para analisar os dados dos portadores de DM do Programa Doce Desafio com a finalidade de extrair conhecimento e informação. Para tanto foram aplicadas na base de dados as técnicas de Estatística, Associação e Clusterização. Os resultados evidenciam a necessidade do planejamento, avaliação e continuidade das ações educativas. Nessa linha, o trabalho apresentou resultados que podem ser facilmente interpretados e ajudam no entendimento de características de nosso público alvo, contribuindo assim para a gestão e organização de campanhas dirigidas a esses perfis promovendo a saúde da população.

Palavras Chaves: Diabetes Mellitus, Mineração de Dados, Educação em Saúde, Exercício Físico.

ABSTRACT

EDUCATION PROGRAM ON DIABETES WITH EMPHASIS ON PHYSICAL EXERCISES: ANALYSIS AND DATA MINING

Author: Thiane Marques Torquato

Supervisor: Prof(a). Dr(a). Prof^ª. Dra. Suézia de Siqueira Rodrigues Fleury Rosa

Co-supervisor: Dr(a). Prof^ª. Dra. Juliana de Faria Fracon e Romão

**Post-Graduation Program in Biomedical Engineering –
Brasília, may of 2016.**

Diabetes Mellitus (DM) is a metabolic disease with multifactorial causes directly related to alterations in the pancreas due to the inability of secreting insulin (*Diabetes Mellitus* type 1 – DM1) and/or producing insulin (*Diabetes Mellitus* type 2 – DM2). The main way to significantly reduce the complications of DM is adhering to diet orientations, to exercise, to drug therapy and to glycemic self-monitoring supported by health education programs. The Program *Doce Desafio* (Sweet Challenge) (PDD) is a health education program of the University of Brasília (UnB) which began in 2001 and has been carried out continuously, aiming to reach an improvement in the quality of life of diabetic people, through oriented physical exercise practice. The patients are monitored in Physical Education classes by filling out daily forms, and consequently, a great amount of data has been collected and stored. Thus, to aid in the investigation of such materials, this study aims to use data mining tools to analyze data from the carriers of DM in the Program *Doce Desafio*, with the purpose of extracting knowledge and information. In order to do so, the techniques of Statistics, Association and Clustering were applied in the database. The results evidence the need of planning, assessment and continuity of the educational actions. In this line, the work showed results which can be easily interpreted and help to understand the characteristics of our target audience, thus contributing to the management and organization of campaigns which are directed to those profiles, promoting general population health.

Key words: Diabetes Mellitus, Data Mining, Health Education, Physical Exercise.

SUMÁRIO

DEDICATÓRIA.....	4
AGRADECIMENTOS	5
RESUMO	7
ABSTRACT	8
SUMÁRIO.....	9
LISTA DE ABREVIACÕES E SIGLAS	11
LISTA DE TABELAS	13
LISTA DE QUADROS	14
LISTA DE FIGURAS	15
1. INTRODUÇÃO	16
1.1 CONTEXTUALIZAÇÃO E FORMULAÇÃO DO PROBLEMA	16
1.2 OBJETIVOS	18
1.2.1 OBJETIVO GERAL.....	18
1.2.2 OBJETIVO ESPECÍFICO.....	18
1.3 JUSTIFICATIVA	19
1.4 ESTRUTURA DA DISSERTAÇÃO.....	21
2 FUNDAMENTAÇÃO TEÓRICA	22
2.1 CENÁRIO DOS PROGRAMAS DE EDUCAÇÃO EM SAÚDE EM DM	22
2.2 EXERCÍCIO FÍSICO E DIABETES MELLITUS	25
2.3 MINERAÇÃO DE DADOS NO CAMPO DA SAÚDE	28
2.4 KDD - DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS	32
2.5 MINERAÇÃO DE DADOS	33
2.6 ANÁLISE DE ASSOCIAÇÃO	39
2.7 CLUSTERIZAÇÃO OU AGRUPAMENTO	41
2.8 ALGORITMO PAM.....	44
2.9 GRÁFICO DA SILHUETA.....	45
2.10 COEFICIENTE DE GOWER.....	47
3 MATERIAIS E MÉTODOS	49
4 RESULTADOS E DISCUSSÃO	54

4.1	PERFIL SOCIODEMOGRÁFICO DOS PARTICIPANTES DO PROGRAMA.....	54
4.2	CARACTERÍSTICAS DOS DESISTENTES DO PROGRAMA.....	63
4.3	INFLUÊNCIA DO TRATAMENTO NA GLICEMIA E PRESSÃO.....	68
4.4	ANÁLISE DA INFLUÊNCIA NA VARIAÇÃO DE GLICEMIA.....	70
4.5	APLICAÇÃO DAS REGRAS DE ASSOCIAÇÃO.....	73
4.6	ANÁLISE DE AGRUPAMENTOS (CLUSTERIZAÇÃO).....	77
5	CONCLUSÃO	81
6	REFERÊNCIAS.....	83
	ANEXO A – PUBLICAÇÕES	96

LISTA DE ABREVIACOES E SIGLAS

ADA - American Diabetes Association

ADA - American Diabetes Association

CO - Centro Olmpico

CP - Convenio Particular

DAC - Doena Arterial Coronariana

DCNT - Doenas Crnicas No Transmissveis

DM - *Diabetes Mellitus*

EASD - European Association for the Study of Diabetes

FC - Fundamental Completo

FI - Fundamental Incompleto

HIPERDIA – Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabticos

ICS - Instituto Curitiba de Sade

KDD - *Knowledge Discovery in Databases* - Processo de Extrao de Informaes de Bases de Dados.

MC- Mdio Completo

OMS - Organizao Mundial da Sade

PAD - Presso Arterial Diastlica

PAM - Partitioning Around Medoids

PAS - Presso Arterial Sistlica

PDD – Programa Doce Desafio

PU – Convenio Pblico

QV - Qualidade de Vida

QVRS -Qualidade de vida relacionada  sade

SAM - Samambaia

SBD - Sociedade Brasileira de Diabetes

SC - Superior Completo

SINAN - Sistema de Informação de Agravos e Notificações

SOB- Sobradinho

LISTA DE TABELAS

Tabela 1: Valores da Silhueta média	47
Tabela 2: Quantidade de Participantes por local de atendimento.....	54
Tabela 3: Quantidade de participantes por local de atendimento e gênero.	55
Tabela 4: Proporção de participantes por local de atendimento e gênero	55
Tabela 5: Idade média por local de atendimento	56
Tabela 6: Idade média de ingresso por local de atendimento.....	57
Tabela 7: Tempo do Diagnóstico ao Ingresso	58
Tabela 8: Quantidade de pacientes por tipo de Diabetes.....	59
Tabela 9: Porcentagem do tipo de diabetes por local de atendimento.....	59
Tabela 10: Quantidade de tipo de consulta por local de atendimento	60
Tabela 11: Proporção do tipo de consulta por local de atendimento.....	60
Tabela 12: Quantidade de escolaridade por local de atendimento	61
Tabela 13: Proporção de escolaridade por local de atendimento	61
Tabela 14: Quantidade de desistentes por local de atendimento	63
Tabela 15: Quantidade de desistentes por sistema de saúde	64
Tabela 16: Proporção de desistentes por sistema de saúde	65
Tabela 17: Proporção de desistentes por escolaridade	66
Tabela 18: Média de presenças por desistência e local de atendimento.....	66
Tabela 19: Estatísticas das variáveis Pressão Diastólica e Sistólica	68
Tabela 20: Teste de hipóteses para as variáveis de interesse.	69
Tabela 21: Estatísticas da variável Glicemia.....	69
Tabela 22: Teste de hipóteses para as variáveis de interesse	70
Tabela 23: Análise de regressão das variáveis de interesse na variação média de glicemia. ...	71
Tabela 24: Associações das variáveis no Cadastro	74
Tabela 25: Associações das variáveis Sistema de Saúde vs Local de Atendimento	75
Tabela 26: Associações das variáveis Insulinoterapia ao entrar vs Tipo de Diabetes Melitius	75
Tabela 27: Associações das variáveis Tipo de Diabetes Melitius vs Local de Atendimento...	76
Tabela 28: Associações das variáveis Gênero vs Local de Atendimento.....	76
Tabela 29: Indivíduos representativos (medóides) em cada grupo	81

LISTA DE QUADROS

Quadro 1 - Resumo das tarefas de Mineração de Dados	35
Quadro 2 - Principais softwares no mercado	38
Quadro 3 - Comparativo de algoritmos usados na análise de agrupamentos	42

LISTA DE FIGURAS

Figura 1 - Processo KDD	32
Figura 2 - Principais Componentes do KDD e DM e seus relacionamentos.....	34
Figura 3 - Tarefas de Mineração de Dados.....	35
Figura 4 - Dados para Mineração em Excel	51
Figura 5: Idade por local de atendimento	56
Figura 6: Idade de ingresso por local de atendimento	57
Figura 7: Tempo de diagnóstico ao ingresso por local de atendimento.	58
Figura 8: Porcentagem dos tipos de DM por local	59
Figura 9: Porcentagem da escolaridade por local.....	62
Figura 10: Porcentagem de desistentes por local.....	63
Figura 11: Porcentagem de desistentes por sistema de saúde	65
Figura 12: Presenças por desistência	67
Figura 13: Diagrama de Dispersão da Variação de Glicemia pela quantidade de Presenças...	71
Figura 14: Diagrama de Dispersão da Variação de Glicemia pela Idade.....	71
Figura 15: Confiança e Suporte das regras de Associação	74
Figura 16: Resultados de silhueta média para agrupamentos com 2 a 10 clusters com o conjunto de 9 variáveis	78
Figura 17: Resultados de silhueta média para agrupamentos com 2 a 10 clusters com o conjunto de 5 variáveis.....	79
Figura 18: Gráfico da silhueta para o agrupamento com 3 clusters.	80

1. INTRODUÇÃO

Neste capítulo serão apresentadas a contextualização e formulação do problema de pesquisa, seguida do objetivo geral e objetivos específicos, bem como a justificativa do estudo.

1.1 CONTEXTUALIZAÇÃO E FORMULAÇÃO DO PROBLEMA

O Diabetes *Mellitus* (DM) é uma doença endócrina, com causas multifatoriais, decorrente da alteração no pâncreas devido a sua incapacidade de produzir insulina, resistência a sua ação ou de ambas (ADA, 2015). O aumento da incidência e prevalência do DM pelo mundo é consequência do estilo de vida pouco saudável da população como dieta inadequada, falta de atividade física e obesidade.

Neste contexto a estimativa global mais recente aponta a existência de 12 milhões de diabéticos no Brasil dos quais 90% são do tipo 2, já no mundo existem 387 milhões de casos de diabetes e espera-se chegar ao total de 592 milhões no ano de 2035 (ATLAS - IDF, 2014).

A prática de exercício físico é recomendada a todos os pacientes para manter o controle glicêmico, pois é um excelente método existente para aumentar a eficácia da insulina, é de fundamental importância também mudança de hábitos alimentares com dieta equilibrada, seguida do monitoramento constante da glicemia capilar além da cessação do tabagismo (ADA, 2015; SBD, 2014).

Dessa forma, com a realização do controle metabólico o paciente mantém-se assintomático e previne-se das complicações agudas e crônicas promovendo a qualidade de vida e reduzindo a morbidade e mortalidade. Vários estudos indicam que os pacientes sabem o quanto benéfico a atividade física é para o controle do Diabetes, mas mesmo assim existe uma baixa adesão a essa prática (COSTA et al., 2011).

Frente a isso as ações educativas para a saúde são a base teórica e metodológica para as ações de promoção da saúde, pois assumem um papel importante nas reduções das complicações e melhoria da qualidade de vida dos indivíduos e ou grupos ao longo do processo da doença, seja tanto pela prevenção quanto reabilitação de doenças. Assim, a educação para saúde passa a ser considerado um processo social relacionada à implementação

de ações pontuais dos profissionais de saúde que incentivem a participação ativa do educando ao processo educativo (REGO et al., 2006).

No entanto, para que as ações educativas gerem aprendizagem é preciso que se pautem em um tipo de educação para saúde acessível, qual seja o modelo dialógico de educação para a saúde que prime pela problematização, construção de conhecimentos e competências, educação baseada no diálogo, mudanças prolongadas de comportamento e maior autonomia ao indivíduo, desencorajando a acomodação e estimulando o auto-cuidado, visando à mudança de uma realidade passível de ser mudada por eles (GROSSI, 2009; FIGUEIREDO et. al,2010).

Mediante o exposto as ferramentas de tecnologia da informação apresentam um grande potencial de contribuição para um melhor gerenciamento dos serviços de saúde, auxiliando os profissionais no diagnóstico e na tomada de decisões. Devido à grande quantidade de informações existentes nos serviços de saúde, a tecnologia da informação vem demonstrando sua utilidade com a nova geração de teorias e ferramentas computacionais (MS, 2009).

O emprego das técnicas de Mineração de Dados permite entender o comportamento dos dados armazenados e permite também a identificação das afinidades existentes entre esses dados (LEITE, 2002). Nesse sentido, a identificação do perfil do paciente com DM no Brasil é um dos primeiros passos para o direcionamento das ações em saúde com o intuito de desenvolver programas eficazes e viáveis para a prevenção do DM com o objetivo de diminuir riscos de complicações, controle de incidências, além de diminuir o impacto financeiro da doença nos cofres públicos (GOMES et.al., 2006).

Atualmente o aumento da análise de informações em grandes bases de dados é devido à velocidade das mudanças e a necessidade de adequação as estas, exigindo a utilização de técnicas e ferramentas que auxiliem na atividade de coleta, análise e processamento desses dados transformando-os em conhecimento útil de forma inteligente e automática. Para Lucas (2002) a transformação da informação em conhecimento pode trazer benefícios às organizações, pois as informações depois de analisadas de forma correta podem ser utilizadas para tomada de decisões mais seguras, aliadas a adequação da postura estratégica da organização.

A comum prática de somente coletar e armazenar os dados, não trás melhoria na estratégia das organizações. É necessário analisar esses dados transformando os em informações com a intenção de descobrir padrões de comportamento que estejam implícitos nessas bases, assim, tornam-se condição primordial na tomada de decisão. Chiara (2003) ressalta que para a aplicação de técnicas de Mineração de Dados, é necessário que se tenha uma coleção de dados disponível. Entretanto, o problema é conseguir dados relevantes para se extrair deles conhecimento potencialmente útil.

Diante do exposto, a contribuição desta dissertação de mestrado refere-se à aplicação de técnicas de Data Mining em ambientes de Educação em Saúde, com a finalidade de extrair conhecimento e informação. Tal análise será feita através da metodologia Mineração de Dados que é formada por um conjunto de ferramentas estatísticas, o processo completo é chamado *Knowledge Discovery in Databases* (KDD), a partir dessa análise pode-se pensar em ações que possam beneficiar esse sujeito quanto ao tratamento e manejo do DM.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Analisar por meio da aplicação de técnicas de mineração de dados, a base de dados de um programa de educação em diabetes com ênfase em exercícios físicos, com a finalidade de extrair conhecimento e informação.

1.2.2 Objetivo Específico

- Analisar as características sociodemográficas e clínicas dos indivíduos atendidos e dos desistentes.
- Analisar a influência de alguns fatores sociodemográficos na variação da glicemia.
- Analisar a influência do acompanhamento do programa Doce Desafio através do exercício físico na melhoria do controle glicêmico e da pressão arterial.
- Aplicar nas bases de dados do Programa Doce Desafio, uma ferramenta de Mineração de Dados na extração da informação e conhecimento.
- Discutir os resultados obtidos.

1.3 JUSTIFICATIVA

As Doenças Crônicas Não-Transmissíveis (DCNT) assumem papel de destaque entre as principais causas de morbimortalidade. Dentre elas o Diabetes *Mellitus* (DM) recebe destaque, pois o seu quadro epidemiológico é alarmante visto que, no Brasil, estudos revelam que 13,3 milhões de pessoas são portadores dessa patologia, uma prevalência de 10,5% na população adulta, sendo o 4º país com maior volume de pessoas diagnosticadas com DM (IDF, 2014).

Vários pesquisadores têm enfatizado que a taxa glicêmica alta é a variável que está envolvida na etiologia das complicações crônicas do DM e, por isso, o controle glicêmico é a principal meta para sua prevenção (SBD, 2009; GUZMÁN et al., 2010). Ações estratégicas para o enfrentamento das DCNT no Brasil 2011-2022, com base nos trabalhos de LI et al. (2010) e da Organização Mundial da Saúde (WHO, 2010) corroboram este enfoque ao assumir que “o controle da glicemia, da pressão arterial e os cuidados com os pés são intervenções plausíveis e custo efetivas para as pessoas com diabetes” (MS, 2011).

O tratamento do Diabetes é muito complexo e requer mudanças comportamentais, os indivíduos em risco de desenvolver DM tipo 2, devem ser estimulados a mudanças no estilo de vida a partir de programas educativos baseados em perda moderada de peso corporal, prática semanal de atividade física (150 minutos/semana), monitoração da glicemia, administração de medicamentos e/ou insulina, adoção de uma alimentação saudável para manter os níveis glicêmicos estabilizados e assim prevenir as complicações crônicas, além de cuidados com os pés para evitar perda de sensibilidade por neuropatia e agravo de ferimentos (SBD, 2013-2014).

Nesse sentido a introdução de tais mudanças depende de uma educação em saúde planejada e estruturada de modo adequado para que os resultados sejam alcançados e principalmente mantidos. Dessa forma, para que se tornem eficazes os programas de educação em saúde, é necessário que estejam voltadas para a realidade para qual se destina, com isso, deve concentrar-se na melhora da saúde e condições de vida da população (RIBAS et.al, 2009).

Em função da alta prevalência do DM estudos têm sido realizados na tentativa de produzir conhecimentos consistentes para minimizar esta problemática e investigar o

benefício da educação no manejo do Diabetes (PIMAZONI NETTO, 2008; MORAES, ROLIM e COSTA JR, 2009). Vários autores defendem a educação do paciente como fundamental para o controle do diabetes e têm recebido bastante atenção em pesquisas nacionais e internacionais que propõem diretrizes (American Diabetes Association, 2010; Funnell et al., 2010; Pimazoni Netto, 2008;; Sociedade Brasileira de Diabetes, 2007)

Contudo, a associação da saúde à tecnologia torna-se uma necessidade, porém, o uso eficiente dos dados e informações como componentes para a tomada de decisão operacional e gerencial na área da saúde ainda é incipiente, se comparada a outros setores (SHAMA et al, 2014).

O descobrimento de conhecimento a partir de bases de dados favorece o surgimento de novos e inusitados modelos de associações de informações, condição impossível de ocorrer num tratamento manual de uma enorme quantidade de dados, como ocorre no Programa Doce Desafio. A informática é uma ciência que colabora com a área da saúde por meio dos sistemas de informações com a finalidade de armazenar e manipular os dados de forma automática (GALVÃO, 2009).

O uso da informática através da aplicabilidade da mineração de dados é benéfico em diferentes áreas da saúde, tendo em vista a quantidade de dados gerados e a importância das informações que podem ser extraídas a partir de prontuários, dados hospitalares, sistemas de informações, entre outros. A Mineração de Dados aplicada nas áreas de Medicina e Saúde Pública permite identificar: terapias, fatores causa/efeito, novas abordagens para os tratamentos e determinar padrões. A análise dos dados de saúde melhora a saúde e melhora o desempenho de tarefas de gerenciamento de pacientes (FAYYAD, 1996).

O cuidado da saúde é um serviço complexo pela sua própria natureza e muito extenso em sua atuação (MS, 2011). Com base nesses pressupostos, esta pesquisa vem contribuir com a ampliação de conhecimento relacionado à temática do gerenciamento de casos, além de evidenciar o benefício da aproximação das ferramentas de tecnologia da informação para as questões relacionadas ao cuidado de pessoas diabéticas.

De acordo com Torres et al. (2010), a educação em DM é o processo para fornecer ao usuário conhecimentos e habilidades necessárias ao desempenho do seu próprio cuidado e para promover as mudanças de comportamento necessárias que levam a administrar a doença com sucesso. Em face a este problema, conhecimentos sobre a mazela e a inserção do

paciente no processo saúde-doença como a participação em programas de tratamento, podem contribuir para uma melhor perspectiva da atenção em diabetes, objetivando a integralidade do cuidado em saúde (QUEIROZ et. al., 2009).

A avaliação de ações educativas proporciona conhecimento e melhora nas estratégias utilizadas. Devido à escassez de estudos, pouco se conhece sobre os efeitos causados pela educação em grupos. No Brasil existem poucos estudos que se dedicaram a avaliar os efeitos das práticas educativas. Poucos são os autores brasileiros se dedicam a essa temática (PAIVA et al., 2004; TORRES et al., 2009; OTERO et al., 2008; TEIXEIRA; ZANETTI, 2006; ZANETTI et al., 2007). Aliada à necessidade de conhecer mais sobre esse assunto, é necessária a produção de conhecimento que crie orientações claras, auxiliem na identificação do perfil do paciente e proporcione subsídios aos profissionais, uma vez que os mesmos reconhecem o seu despreparo para conduzir essas ações (SANTOS, 2011; TORRES et al. 2010).

Diante desta perspectiva, a realização de estudos que investiguem o efeito dos Programas de Educação em Saúde para instalar comportamentos de adesão ao tratamento é essencial para que a prática do profissional de saúde seja baseada em evidências científicas e os gestores serem capacitados a tomar decisões relacionadas aos serviços prestados a população, de acordo com suas reais necessidades (OMS, 2003).

1.4 ESTRUTURA DA DISSERTAÇÃO

O trabalho está organizado como segue: O Capítulo 1, Introdução, apresentou por meio de sua contextualização ao tema proposto neste trabalho. Da mesma forma foram estabelecidos os resultados esperados por meio da definição de seus objetivos gerais e específicos bem como a justificativa da pesquisa.

O segundo capítulo apresenta a Fundamentação Teórica que orienta a investigação, complementada por trabalhos científicos (monografias, dissertações, teses e artigos científicos específicos). Neste capítulo são abordados os assuntos pertinentes ao trabalho como: Cenário dos Programas de Educação em Saúde em DM, a definição de Exercício Físico e Diabetes Mellitus, Pesquisas com o tema Mineração de dados no campo da saúde, Definição de

Mineração de Dados, Análise de Associação, Clusterização ou Agrupamento e Algoritmo PAM.

O Capítulo 3 apresentou-se a Metodologia da pesquisa utilizada. O capítulo 4 traz os resultados, discussão e aplicação das técnicas de Associação e de Clusterização dos dados analisados.

No Capítulo 5, são apresentadas as conclusões do trabalho, relacionando os objetivos identificados inicialmente com os resultados alcançados. São ainda propostas possibilidades de continuação da pesquisa desenvolvida a partir das experiências adquiridas com a execução do trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 CENÁRIO DOS PROGRAMAS DE EDUCAÇÃO EM SAÚDE EM DM

Para promover o controle efetivo da DM e a mudança de comportamentos relacionados à doença, a Sociedade Brasileira de Diabetes propõe a utilização de estratégias educativas no intuito de aumentar a adesão dos usuários, pois é uma tarefa difícil de ser alcançada (VIEIRA, 2011).

A Educação em Diabetes se destaca como um dos pilares do tratamento dessa doença, porém os estudos são heterogêneos quanto aos tipos de intervenção para populações específicas, não havendo um programa universal de educação em diabetes que possa ser padronizado e reconhecido como eficaz em todo o mundo. Existe a orientação para que se realizem práticas educativas nos níveis da atenção básica para os usuários com DM: hipertensos, adolescentes, mulheres, idosos e outros públicos em específico (BRASIL, 2007).

Em suas diretrizes, a SBD propõe que as atividades educativas devam contemplar a participação do usuário no processo educativo; ser parte integrante de todas as atividades; possuir envolvimento multidisciplinar e planejamento prévio com estabelecimento dos seus objetivos a curto, médio e longo prazo, além de avaliar continuamente todo o processo (SBD, 2007).

Em uma pesquisa realizada na Catalunha com uma amostra de 276 pessoas diabéticas, do tipo 2, constatou melhor controle metabólico nas que participam de intervenção educativa especializada (CAMPOS et al., 2009).

Outro trabalho, realizado na Califórnia, foi avaliado em um período de nove meses um total de 336 pacientes com DM tipo 2. Os autores fizeram o acompanhamento telefônico e a estratégia de visita domiciliar, como objetivo de verificar o grau de envolvimento da equipe médica e dos pacientes com o plano de cuidados. Foi constatado que os pacientes que possuem baixa escolaridade se comprometeram mais no programa educativo e apresentaram maior adesão com o programa de visita domiciliar (SCHILLINGER et al., 2007).

Estratégias de educação em grupo permitem o compartilhamento de informações com outras pessoas e a troca das experiências vivenciadas com a doença contribuindo positivamente para maior adesão da pessoa ao tratamento, já que possibilita o apoio mútuo dos participantes. Nesta direção, um estudo comparou as estratégias de educação em grupo e individual no programa educativo de um hospital de Belo Horizonte (MG) e constatou que as duas estratégias do programa educativo em Diabetes foram efetivas, porém a educação em grupo apresentou melhores resultados de controle glicêmico do que a individual. Os autores afirmam que a educação em grupo fortalece o elo entre o profissional de saúde e o indivíduo melhorando a condição clínica (TEIXEIRA et. al., 2005; TORRES et. al. 2009).

Em consonância Rickheim (2002) demonstrou que a educação realizada de maneira individual (por intermédio de consultas com nutricionistas, enfermeiras ou outros educadores) comparada à educação em grupos de pacientes, utilizando-se a mesma metodologia, foi equivalente na melhora do controle metabólico, visto que o aprendizado em grupo apresentou melhor relação custo-benefício.

Já outra pesquisa realizada no Reino Unido, entre 824 pessoas com DM tipo 2, por meio de um estudo do tipo *cluster* com randomização para o programa de educação especializada, buscou verificar a efetividade do programa educativo sobre as variáveis glicohemoglobina, peso e práticas de autocuidado. Foi encontrado maior redução dos valores de glicohemoglobina e peso entre os participantes do grupo de intervenção do que entre aqueles que fizeram parte do grupo controle (DAVIES et al., 2008).

Torres (2004) e Leite et al. (2008) apontam a necessidade de se abordar conteúdos básicos em um programa educativo em DM como a fisiopatologia, noções fundamentais acerca da terapia insulínica, sintomas, aspectos emocionais, gerenciamento de complicações, nutrição, cuidados com os pés, exercício físico, automonitorização, distribuição de material ilustrativo, apresentações de aulas didáticas, até intervenções envolvendo a participação do paciente ativamente.

A principal meta do aprendizado é treinar o portador de Diabetes a tomar decisões efetivas em seu autocuidado, devendo-se considerar que todas as estratégias educativas precisam de continuidade, sendo necessário adaptá-las às necessidades de aprendizagem e experiência de cada usuário e que a aquisição do conhecimento e a adesão às práticas fazem parte de um processo (SILVIA, 2010).

Em adição, Faria et. al. (2013) realizaram um estudo quase-experimental num serviço de atenção primária de um município do interior paulista, em 2008, para avaliar a qualidade de vida relacionada à saúde de pessoas com DM antes e após participação em um programa educativo de cinco meses. Participaram 51 sujeitos com predomínio do sexo feminino (56,9%), média e desvio-padrão de $57,65 \pm 11,44$ anos de idade. Foi constatado que a participação dos sujeitos no programa educativo em DM também contribuiu para a melhoria da percepção acerca de seu estado geral de saúde.

O estímulo de práticas saudáveis deve atingir ações educativas com os idosos e podem contribuir na melhora significativa da sua saúde e autonomia. Tal afirmativa corrobora com uma pesquisa realizada na região Metropolitana de Curitiba que utilizou roda de conversa como estratégia para observar os pacientes acima de 60 anos cadastrados no Hiperdia (Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabéticos) tornando um espaço de acolhimento e troca de experiências e constatou que relato de experiência pode contribuir para o desenvolvimento de pesquisas sobre estratégias de educação em saúde, aumentar a adesão à terapêutica, evitar e/ou diminuir as complicações decorrentes da patologia e melhorar a qualidade de vida dos doentes crônicos (LOPES et. al., 2009).

Mantovani et al. (2011) consideram que as ações educativas possibilitam “maior conhecimento da doença e suas complicações” pelos indivíduos, condições essenciais para atingir o controle da enfermidade. O caminho mais acertado para que os programas e políticas de saúde consigam reconhecer a importância do incentivo e a responsabilidade social em

todas as fases da vida dos diabéticos, é através de medidas protocolares já estabelecidas pelas evidências científicas para o favorecimento do controle dessas DCNT.

Além disso, os programas em saúde devem fornecer condições para o desenvolvimento de habilidade para a tomada de decisão em vários âmbitos sejam eles, biológicos, sociais e psicológicos. Requerem uma proposta de ensino que propicie o aprendizado de forma autônoma e reflexiva respeitando conhecimentos e experiências anteriores (NASCIMENTO et. al., 2010).

Para Leite, Prado e Peres (2010) a função educativa deve envolver não somente o paciente, mas também a família para conscientização desses indivíduos na manutenção da saúde. Para o alcance desse objetivo, certamente será necessária a implementação de práticas pedagógicas inovadoras que valorizem o diálogo e estimulem a reflexão-ação dos indivíduos.

A fundadora e coordenadora do Programa Doce Desafio (PDD), Prof^a Jane Dullius, escreveu o livro “Diabetes *Mellitus*: Saúde, educação e atividade físicas (2007)” relatando detalhes sobre sua vivência com a diabetes. O convívio com a diabetes desde muito cedo reflete nas diretrizes e na forma de trabalho do PDD fundamentada também com o estudo de mestrado e doutorado sobre o tema da DM. O PDD tem a função de viabilizar a autonomia do sujeito portador de DM perante sua própria condição por meio da educação. (DULLIUS, 2003)

Diante disso, os programas de saúde direcionados aos portadores de Diabetes devem basear nas seguintes estratégias de abordagem para motivação dos pacientes: educação, modificação do estilo de vida, reorganização de hábitos alimentares, orientação na prática de atividade física e uso de medicamentos (BARSAGLINI, 2011).

2.2 EXERCÍCIO FÍSICO E DIABETES MELLITUS

Segundo definição do Ministério da Saúde DM é um grupo de doenças metabólicas caracterizadas por hiperglicemia e associadas a complicações como disfunções e insuficiência de vários órgãos, especialmente olhos, rins, nervos, cérebro, coração e vasos sanguíneos. Relacionada diretamente a alterações no pâncreas devido a defeitos de secreção e/ou ação da

insulina envolvendo processos patogênicos específicos, por exemplo, destruição das células beta do pâncreas (produtoras de insulina), distúrbios da secreção da insulina, entre outros (ADA, 2015).

O exercício físico é indicado tanto para pessoas com DM do tipo 1 quanto do tipo 2, desde que a prática seja orientada e realizada com controle médico e profissional. É considerado um grande aliado no tratamento do Diabetes reduzindo o risco cardiovascular, atuando sobre o controle glicêmico, promovendo maior capilarização das fibras musculares, melhorando a função mitocondrial e aumentando a captação da glicose sanguínea para os músculos. Observa-se maior sensibilidade à insulina nas 24 a 72 horas após uma sessão de exercício aumentando a captação da glicose nos músculos e nos a dipócitos e reduzindo a glicemia sanguínea. O Diabetes reduz a expectativa de vida em 5-10 anos e aumenta o risco de doença arterial coronariana (DAC) em 2-4 vezes. O exercício é um forte aliado na predição desse risco nos diabéticos assim como na sua redução por meio da prática regular (SBD 2013-2014).

Nesse sentido, a *American Diabetes Association (ADA)* e a *European Association for the Study of Diabetes (EASD)*, recomendam a prática semanal (150 minutos) de atividade aeróbica de moderada intensidade ou pelo menos 90 minutos de atividade aeróbica que atinja índices acima de 70% da frequência cardíaca máxima, divididos no mínimo em 3 dias da semana. A ADA também inclui, que na ausência de contra indicações, pessoas com DM2 devem ser estimulados a praticar exercícios resistidos por no mínimo três vezes na semana, com o objetivo de trabalhar os principais grupos musculares (SIGAL et. al., 2006).

Corroborando com esse fato, o Caderno de Atenção Básica sobre o DM do Ministério da Saúde, a incidência do *Diabetes Mellitus* foi reduzida 58% em 3 anos somente com a mudança do estilo de vida, recomendações para prevenção de DM2 incluem a manutenção um peso saudável, consumindo uma dieta saudável, e participação no exercício. Essas mudanças visam discreta redução de peso (5-10% do peso), manutenção do peso perdido, aumento da ingestão de fibras, restrição energética moderada, restrição de gorduras, especialmente as saturadas, e aumento de atividade física regular (SBD, 2013-2014).

O exercício é bastante benéfico no aumento da eficácia da insulina e com a prática regular moderada ele se torna um excelente método para melhorar o gasto calórico pelos pacientes diabéticos. Em contrapartida eles tendem a perder facilmente o controle da

estabilidade da insulina, por isso devem manter o equilíbrio entre a quantidade e frequência das refeições com a prática de exercícios moderados regulares. No entanto, os exercícios exaustivos podem provocar hipoglicemia nos pacientes desinformados, por essa razão, os diabéticos devem ser orientados a reduzir a dose de insulina ou ingerir suplementos de carboidratos, antes das atividades físicas (Greenspan e Gardner, 2006).

Segundo Saad, Maciel e Mendonça (2007) a atividade física deve ser compreendida, avaliada e aplicada por uma equipe multiprofissional capaz de analisar os riscos e benefícios do exercício para cada paciente. Ressaltam ainda que a prática da atividade física deve fazer parte do tratamento, ou seja, um instrumento dentro da terapêutica. No entanto, a Sociedade Brasileira de Diabetes salienta que estabelecer um protocolo preciso para aplicação do exercício físico é difícil, pois a resposta metabólica ao exercício depende de fatores específicos de cada indivíduo, por exemplo, tipo de exercício, intensidade, horário, duração, tipo de refeição, horário da última dose de insulina, controle metabólico complicações e ainda nas mulheres a fase do ciclo menstrual (SBD, 2009).

Apesar das inúmeras vantagens da atividade física, a Sociedade Brasileira de Cardiologia afirma que grande parte da população não possui o hábito de praticar exercícios, sendo considerada inativa, ou ainda se exercitam muito pouco, em níveis insuficientes para obter resultados satisfatórios para a saúde, mostrando que o sedentarismo é um problema que vem crescendo significativamente no Brasil. Estima-se também que 50% dos indivíduos que começam a participar de um programa de exercícios podem interrompê-lo nos primeiros seis meses (SBC, 2010).

Mesmo sabendo que a atividade física é essencial para o controle e qualidade de vida dos pacientes diabéticos estudos indicam que há baixa adesão a essa prática, sinalizando para a necessidade do desenvolvimento de estratégias e intervenções que levem o indivíduo a motivar-se e assumir uma atitude mais ativa para o controle de sua doença, gerando mudanças comportamentais eficazes com conseqüentes melhoras nas características clínicas (COSTA et. al., 2011; MODENEZE, 2011).

Estudos indicam que apenas 19 a 30% dos usuários que possuem DM aderem às prescrições de exercícios, bem como possuem dificuldades para sua prática diária (NEGRI et al., 2010; GLEESON-KREIG, 2008 ; MILLER et al., 2010). Corroborando com esse fato, Vancea et al. (2009) relatam que a baixa adesão ao exercício físico por diabéticos pode ser

decorrente da falta de estímulos nas consultas de rotina e da falta de compreensão e aceitação dos próprios doentes. Conforme o exposto, podemos destacar a eficácia dos programas em educação e comprovar que existe relação entre qualidade de vida e adesão ao tratamento.

2.3 MINERAÇÃO DE DADOS NO CAMPO DA SAÚDE

Vários pesquisadores têm utilizado a Mineração de Dados para extrair conhecimento útil em grandes bases de dados (FAYYAD, 1996), com o objetivo de explicar a informação e a realizar previsões a partir da mesma (WITTEN et.al., 2011). Neste contexto, as áreas que mais aplicam essa tecnologia é a área de ciências biológicas e da saúde, devido à expansão da área de informática em saúde. Seguindo esse contexto, as pesquisas foram direcionadas para mineração e a identificação de estratégias para a efetividade no uso dos padrões descobertos para a área da Saúde. Os estudos encontrados apóiam essa abordagem, pois geram fortes evidências sobre a eficiência da Mineração de Dados na descoberta de conhecimentos, proporcionando eficiência, melhores tratamentos e ganho de qualidade de vida para os pacientes.

Na área da ciência da saúde, um estudo financiado pela Organização Mundial de saúde (OMS), aplicou a Mineração de dados em grandes e diversas bases de dados para investigar o relacionamento da taxa de prevalência de HIV/AIDS com os grupos de fatores (sociais, econômicos e mão de obra nos cuidados de saúde). Este estudo demonstrou a necessidade de investimento para promover a educação de enfermeiros para os programas de HIV/AIDS e conseqüentemente atingir um impacto positivo na taxa de prevalência. (MADIGAN et. al., 2008).

Por meio de um estudo de caso na base de dados de uma empresa-plano de saúde, Leite (2002) apresentou em seu trabalho uma análise em relação ao processo de descoberta de conhecimento para obter informações relevantes acerca de gastos, consultas, exames e direcionamentos, onde foi possível estabelecer uma relação entre o sexo do usuário, plano do usuário, estado civil e data de nascimento, classificando determinados grupos, através de características semelhantes e ambientando todas as regras e padrões encontrados com a realidade local, que certamente podem servir como indicativos para a direção da empresa.

Trindade et al. (2012) utilizaram a base de dados do SINAN (Sistema de Informação de Agravos e Notificações) do Sistema Único de Saúde para identificar padrões de comportamento das Hepatites Virais. Foram analisadas sessenta e cinco características dos 3.063 formulários de investigação, resultando em quatro árvores de decisão e 99 regras de classificação. Destas regras, 60 foram analisados e os outros foram descartados por não contemplarem exemplos suficientes para serem considerados válidos ou porque tinham um elevado número de erros. Essa análise pôde ajudar na formulação de novas políticas o que sugere a importância deste método como uma forma de aumento de estratégias, com o objetivo de reduzir e controlar doenças.

Já Dallagassa (2009) propôs um estudo quantitativo, retrospectivo, de cunho descritivo e utilizou a mineração de dados através do algoritmo Árvores de Decisão C4.5 a uma base de dados de beneficiários de uma operadora privada de saúde por meio das informações dos procedimentos realizados, com o objetivo de identificar pessoas com indicativos para o Diabetes *Mellitus* Tipo 2, para o ingresso em programas de promoção da saúde e prevenção de doenças, programas estes, que se apresentam mais eficientes na redução de custos para a operadora e na melhoria da qualidade de vida para os beneficiários, quando utilizados no conceito de prevenção primária, ou seja, antes da manifestação da doença. As experimentações realizadas validam a metodologia para a detecção precoce de pacientes com propensão a Diabetes *Mellitus* tipo 2, como base o protocolo de diagnóstico deste mal.

No trabalho de Vianna (2010) são apresentadas tarefas de classificação utilizando o programa WEKA, para identificar características materno-fetais na predição da mortalidade infantil. Com a análise das regras geradas pela mineração de dados, obteve-se o perfil da mortalidade infantil no Estado do Paraná, no período de 2000 a 2004, destacou-se que 55% dos óbitos poderiam ser evitáveis se houvesse uma adequada atenção à gestação, ao parto e ao recém-nascido. Todo o procedimento foi baseado nos protocolos de saúde infantil e apoiado por especialistas em Saúde para validação dos padrões encontrados.

No tocante de identificar associações que indiquem padrões de risco e que possibilite a indicação de usuários com doenças cardiovasculares para o ingresso em programas de gerenciamento de casos, Kobus (2006) propôs um modelo através do algoritmo Apriori para identificar padrões em base de dados de usuários do Instituto Curitiba de Saúde (ICS), onde foi constatado que os usuários incluídos na base de dados apresentaram perfil de elevada

morbidade relacionada a doenças cardiovasculares, sendo indicados para um programa de gerenciamento de casos de doenças cardiovasculares.

Steiner et al. (2006) organizaram um estudo em uma base contendo 118 pacientes que apresentavam quadro de Icterícia por câncer ou cálculo biliares. Seus experimentos se basearam no uso de técnicas de análise exploratória dos dados, realizando o pré-processamento para diferentes tarefas de Mineração de Dados, visando obter aquela que discriminasse padrões com a máxima precisão, auxiliando à tomada de decisão dos especialistas médicos quanto aos seus diagnósticos.

Já Stein Junior (2008) utilizou o KDD através do algoritmo J48 em uma base de dados de micro regiões da cidade de Curitiba, com objetivo de especificar um sistema de informações para monitorar condições de micro áreas urbanas para planejamento de ações para melhoria da qualidade de vida dessa população. Ele ressalta que o motivo das ausências de novos padrões foi devido aos atributos determinados pelos especialistas entrevistados. A tomada de decisão é apoiada por regras extraídas de um protocolo prévio. Por exemplo, regras que associavam condições de precariedade da urbanização de uma micro área com baixo risco para a população foram descartadas pelos especialistas, justamente porque o critério de avaliação adotado determina que uma área somente pode ser considerada de baixo risco se não apresentar precariedades.

Minnie (2011) realizou uma pesquisa no laboratório de patologia clínica para analisar as funções de um contador automático de células sanguíneas. Para melhorar a qualidade do sistema aplicou a fase de pré-processamento do KDD (*Knowledge Discovery in Databases*) realizando a limpeza dos dados, seleção dos dados e transformação de dados do contador de células sanguíneas.

Moura (2007) através da ferramenta de mineração de dados *Spotfire* analisou os dados de pacientes que foram submetidos a exames de glicemia em um período de seis anos, esses dados foram preparados e avaliados seguindo as etapas do processo de KDD. Através dos padrões encontrados foi possível ilustrar as idades que estão realmente procurando tratamento adequado.

Nos Estados Unidos os pesquisadores da Universidade da Johns Hopkins verificaram que as tecnologias de MD podem ser utilizadas em meta-análise para estudar acidentes de trânsito de pessoas idosas. (BAYAM et. al., 2005).

Outro estudo realizado na Índia investigou crianças com Diabetes *Mellitus*, por meio da aplicação da MD e da tarefa de classificação e conseguiram identificar de forma mais rápida através dos atributos do exame físico e bioquímicos laboratoriais novos pacientes que necessitava começar o tratamento, diminuindo assim o sofrimento das crianças. (KAUR e WASAN, 2006).

Já na Alemanha, Pfaff M. et al., (2004) investigaram qual o risco de doenças cardiovasculares em pacientes em hemodiálise, através da tecnologia MD juntamente com a tarefa de clusterização. Eles validaram um método para apoio da tomada de decisão com *clusters* para prever o risco individual dos pacientes em hemodiálise e, e obtiveram um excelente sucesso.

Em Criciúma- SC, os pesquisadores aplicaram a MD e utilizaram um *software* Orion *Data Mining*, ferramenta gratuita desenvolvida pela Universidade do Extremo Sul Catarinense-UNESC para realizar um estudo sobre alergia e rinite em escolares, com objetivo de verificar a prevalência destas causas numa base de dados existentes (MATTOS et. al., 2006)

Podemos perceber com os exemplos acima que em diversos países está crescendo gradativamente os estudos realizados através da MD nas áreas de ciências biológicas, e também subárea da genética, devido à imensidão de dados gerada por essas áreas, esses estudos evidencia a importância da Mineração de Dados no campo da saúde, fornecendo recursos para realização de tomada de decisão criteriosa, promovendo melhores condições para planejamento da utilização e implantação dos recursos de saúde no país (GALVÃO, 2009).

A partir desses exemplos podemos perceber que existem inúmeras técnicas e algoritmos que podem ser utilizadas para extrair informações relevantes. Segundo Fayyad (1996) não existe um modelo universal capaz de resolver todos os problemas. A escolha da técnica para resolver uma situação particular deve ser considerada com atenção, é necessário

definir o problema e quais metas desejam ser alcançadas a fim de obter melhor resposta ao problema.

2.4 KDD - DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

O processo de descoberta de Conhecimentos em Base de Dados (KDD - *Knowledge Discovery in Databases*) foi definido por Fayyad et al. (1996) como: um processo não trivial para identificar padrões que sejam válidos e previamente desconhecidos, potencialmente úteis e compreensíveis, visando melhorar o entendimento do problema, possibilitando o auxílio no processo de tomada de decisões. Ele pode ser visto como o processo de descoberta de padrões e tendências por análise de grandes conjuntos de dados, tendo como principal etapa o processo de mineração.

Este processo pode ser aplicado em diversas áreas entre elas: mercado, análise de crédito, análise de tendências, marketing, astronomia e medicina. Na medicina são encontrados diversos trabalhos aplicados a análise do histórico de pacientes para aperfeiçoar o tratamento e auxiliar na obtenção de conhecimento implícito, para sugerir ações na área da saúde (VIANNA, 2007; KOBUS, 2006).

A Figura 1 ilustra um exemplo proposto por Fayyad (1996) que consiste em cinco passos básicos: Seleção, Pré-Processamento, Transformação, Mineração de dados e Interpretação.

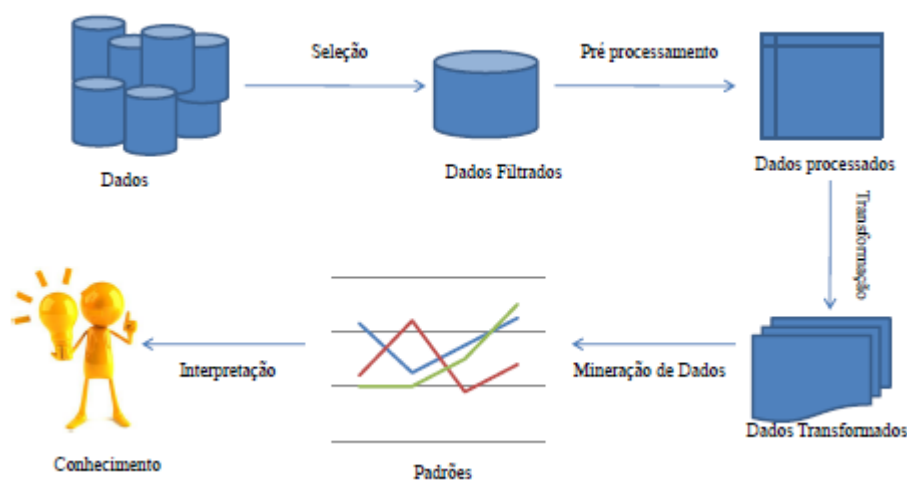


Figura 1 - Processo KDD

Fonte: Adaptado pela autora de (FAYYAD, 1996).

Segundo Boente, Goldschmidt e Estrela (2008) o processo de Descoberta de Conhecimento envolve várias etapas e tem início na análise do entendimento do domínio da aplicação e dos objetivos a serem realizados. Após a fase inicial o foco passa a ser a escolha ou seleção da massa de dados a ser minerada, podendo ser um conjunto de dados ou um subconjunto de variáveis onde a extração será realizada. Tem por objetivo assegurar a qualidade dos dados envolvidos no KDD realizando operações básicas como a remoção de ruídos, que podem ser, por exemplo, atributos nulos. Nesta fase é necessário efetuar a identificação de quais informações da base de dados existentes devem ser efetivamente consideradas durante o processo de KDD.

A segunda fase consiste na Seleção e Transformação dos dados em que serão selecionados os atributos interessantes ao usuário, além de transformados utilizando o padrão ideal para aplicar algoritmos de mineração. Após a realização das fases anteriores a Mineração de Dados (*Data Mining*) é iniciada. Esta fase é a mais importante do KDD sendo realizada através da escolha do método e do algoritmo mais compatível com o objetivo da extração, a fim de encontrar padrões nos dados que sirvam de subsídios para descobrir conhecimentos ocultos.

A Avaliação ou Pós-Processamento é a fase que identifica, entre os padrões extraídos na etapa de *Data Mining*, os padrões interessantes ao critério estabelecido pelo usuário, podendo voltar à fase inicial para novas iterações. Ao término da avaliação o conhecimento descoberto deverá ser implantado e incorporado ao sistema, sempre documentando e publicando os métodos, a fim de apresentar o conhecimento descoberto ao usuário.

2.5 MINERAÇÃO DE DADOS

Diante do avanço das tecnologias no tratamento de bases de dados das organizações não tem sido difícil o armazenamento nem o resgate, quando necessário, de grandes volumes de dados em computadores. Mas, apesar dos dados armazenados serem um bem valioso de uma organização muitas se deparam com o problema de ter muitos dados, mas pouco conhecimento. Tendo em vista a grande quantidade de dados armazenados se torna difícil interpretá-los sem o auxílio de técnicas que permitam a sua automatização e análise de forma inteligente. As técnicas que buscam transformar os dados armazenados em conhecimento são o objetivo da área chamada de KDD (FAYYAD, 1996). O KDD refere-se a todo processo de

descoberta de conhecimento útil em bases de dados, enquanto *Data Mining*, principal etapa do KDD, refere-se à aplicação de algoritmos para extrair modelos dos dados. A Figura 2 ilustra o processo de descoberta de conhecimento e de mineração de dados.

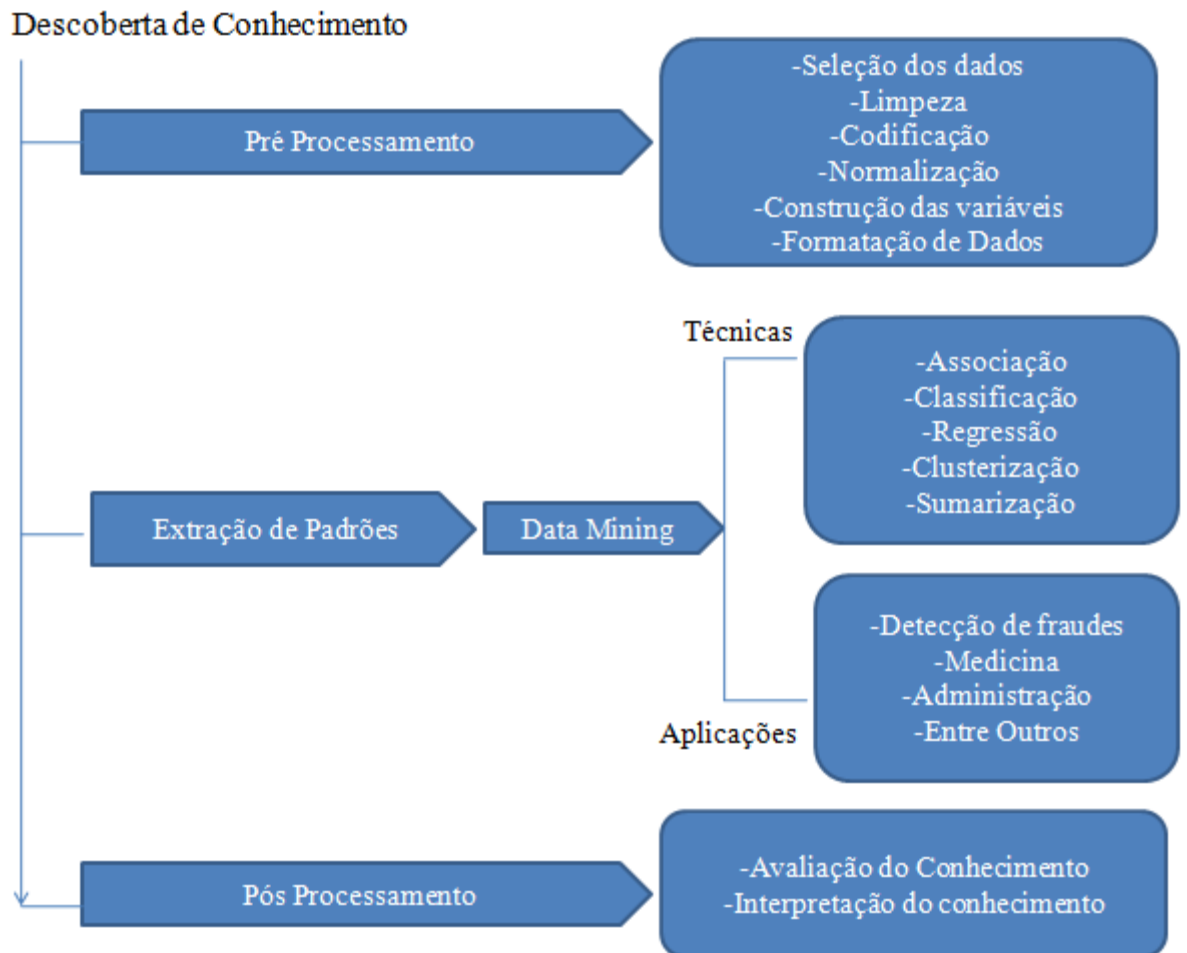


Figura 2 - Principais Componentes do KDD e DM e seus relacionamentos

Fonte: Adaptado de LEE et. al. (2000).

De acordo com Fayad et al. existem dois tipos de objetivos no processo de extração de conhecimento: verificação e descoberta. Na verificação o sistema se limita a verificar as hipóteses definidas pelos usuários enquanto na descoberta o sistema automaticamente encontra novos padrões nos dados. O objetivo de descoberta pode ser dividido em dois outros objetivos: predição e descrição, como mostra a figura 3. Predição o sistema encontra padrões para prever o comportamento futuro de algumas entidades e na descrição o sistema encontra padrões para apresentar aos usuários de forma compreensível por humanos.

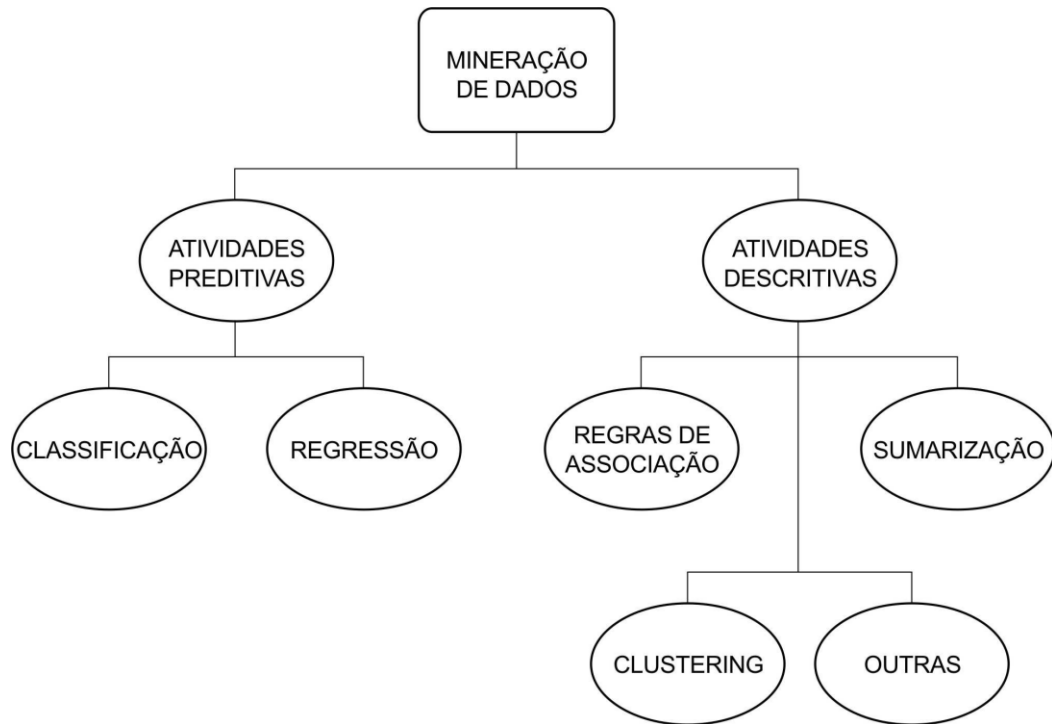


Figura 3 - Tarefas de Mineração de Dados

Fonte: Resende (2005)

Existem diversas técnicas de mineração de dados: classificação, regressão, clusterização, sumarização, associação, entre outras. Para cada uma dessas técnicas já existem diversos algoritmos desenvolvidos.

Dias (2001) cita as principais tarefas de Mineração de Dados, suas descrições e exemplifica-as, conforme pode ser visto no Quadro 1:

Quadro 1 - Resumo das tarefas de Mineração de Dados

Tarefa	Descrição	Exemplos	Tecnologia -aplicativos
Classificação	Constrói um modelo de algum tipo que possa ser aplicado a dados não classificados a fim de categorizá-los em classes.	Classificar pedidos de crédito; Esclarecer pedidos de seguros fraudulentos;	Tecnologia: Redes Neurais, Algoritmos Genéticos, Lógica Indutiva Aplicativos: Pollyanalyst, Xpert/Rule Miner, Microsoft Data Analyzer 2002, Oracle 9i Data Mining, Darwin, Mineset, WEKA, Intelligent Miner, See5, Clementine, Data-Miner Software Kit, MLC++
Estimativa	Usada para definir um	Estimar o número de	Tecnologia: Estatística, Redes

ou Regressão	valor para alguma variável contínua desconhecida.	filhos ou a renda total de uma família; Estimar o valor em tempo de vida de um cliente; Prever a demanda de um consumidor para um novo produto	Neurais Aplicativos: Pollyanalyst, Darwin, Mineset, WEKA, Cubist, Data-Miner Software Kit, MLC++Pollyanalyst, Xpert/Rule Miner, Microsoft Data Analyzer 2002, Oracle 9i Data Mining, Darwin, Mineset, WEKA, Intelligent Miner, See5, Clementine, Data-Miner Software Kit, MLC++
Associação	Estabelece uma correlação estatística entre atributos de dados e conjuntos de dados.	Encontrar relacionamentos ou padrões freqüentes entre conjuntos de dados.	Aplicativos: Apriori, GSP, DHP Pollyanalyst, Magnum Opus, Xpert/Rule Miner, Data Mite, Oracle 9i Data Mining, Mineset, WEKA, Intelligent Miner, R, Clementine
Clusterização	Processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos	Agrupar clientes por região do país; Agrupar clientes com comportamento de compra similar; Agrupar seções de usuários Web para prever comportamento futuro de usuário.	Aplicativos: K-Means, K-Modes, K-Prototypes, Kohonen Pollyanalyst, Xpert/Rule Miner, Microsoft Data Analyzer 2002, Darwin, R, Intelligent Miner Mineset, Clementine, MLC++
Sumarização	Envolve métodos para encontrar uma descrição compacta para um subconjunto de dados.	Tabular o significado e desvios padrão para todos os itens de dados; Derivar regras de síntese.	Tecnologia: Lógica Indutiva, Algoritmos Genéticos Aplicativos: Pollyanalyst, Intelligent Miner

Fonte: Dias (2001) e Goldschmidt e Passos (2005),

A mineração de dados conta com um conjunto de ferramentas, cada ferramenta possui uma abordagem particular em relação ao modo de resolução do problema. Na vasta variedade de conjuntos de dados existentes uma ferramenta de mineração de dados pode num ambiente demonstrar bons resultados, mas em outro ser totalmente inadequada (LAHIRI, 2006). A seguir são apresentadas as principais ferramentas de mineração de dados:

Segundo Camilo e Silva (2009), a classificação é uma das tarefas mais utilizadas na mineração de dados que visa identificar a qual classe um determinado registro pertence, ou seja, entendem os dados existentes ou prevêem como novos dados irão se comportar. Nesta tarefa, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de 'aprender' como classificar um novo registro (aprendizado supervisionado). A tarefa de classificação pode ser usada, por exemplo, para: determinar quando uma transação de cartão de crédito pode ser uma fraude; identificar em

uma escola qual a turma mais indicada para um determinado aluno; diagnosticar onde uma determinada doença pode estar presente; identificar quando uma pessoa pode ser uma ameaça para a segurança; classificação de clientes em baixo, médio ou alto risco de empréstimo bancário; de clientes potencialmente consumidores de um determinado produto a julgar pelo seu perfil.

A regressão é similar à classificação, porém usada quando o registro é identificado por um valor numérico e não um categórico. Assim, pode-se estimar o valor de uma determinada variável analisando-se os valores das demais. Estimar algum índice é determinar seu valor mais provável diante de dados do passado ou de dados de outros índices semelhantes sobre os quais se tem conhecimento. (CAMILO e SILVA 2009).

Para Hamerly e Elkan (2012) a clusterização ou agrupamento é uma tarefa descritiva com o objetivo de identificação de grupos, *clusters*, por meio de agrupamento natural de itens da base de dados, em função da similaridade entre eles. Seu objetivo é formar grupos baseados no princípio de que esses grupos devem ser o mais homogêneos em si e mais heterogêneos entre si. A diferença fundamental entre agrupamento e classificação é que na classificação as classes são pré-definidas pelo pesquisador, enquanto que aqui não existe tal requisito. Isto torna esta técnica muito mais complexa do que a classificação. Na análise de agrupamentos os grupos ou classes são construídos com base na semelhança entre os elementos cabendo ao analisador das classes resultantes avaliarem se estas significam algo útil. Por exemplo, agrupar sintomas pode gerar classes que não representem nenhuma doença explicitamente, uma vez que doenças diferentes podem possuir os mesmos sintomas (HAMERLY e ELKAN, 2012).

Regras de associação é uma das tarefas mais conhecidas devido aos bons resultados obtidos. A tarefa de associação consiste em identificar quais atributos estão relacionados entre si. Apresentam a forma: SE atributo X ENTÃO atributo Y, utilizado principalmente nas análises da "Cestas de Compras" (*Market Basket*) onde identificamos quais produtos são levados juntos pelos consumidores. Entende-se que a presença de um item implica necessariamente na presença do outro na mesma transação (CAMILO e SILVA 2009).

Para possibilitar que profissionais de outras áreas possam fazer uso da Mineração de Dados, segundo Camilo e Silva (2009) diversas ferramentas foram desenvolvidas no intuito

de tornar a aplicação da Mineração de Dados uma tarefa menos técnica. No Quadro 2 é descrito os principais softwares no mercado.

Quadro 2 - Principais softwares no mercado

Nome	Fabricante	Vantagens	Funcionalidade
Clementine	SPSS	Interface orientada a objeto	Suporta o processo CRISP-DM
SAS Enterprise Miner Suite	SAS	Avançada ferramenta para predição e descrição de dados	Possui módulos para trabalhar em todas as etapas do processo de mineração.
WEKA	The University of Waikato	É uma das melhores ferramentas livres	Fornecer as funcionalidades para pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização.
Oracle Data Mining (ODM)	Oracle	Ajuda a transformar gigantes massas de dados em inteligência corporativa	Uso em seu banco de dados ORACLE.
IBM Intelligent Miner	IBM	Poderosa ferramenta para análise de dados integrada	Ferramenta de mineração da IBM para a mineração de dados no banco de dados DB2.
MDR	Ferramenta livre	Ferramenta livre	Ferramenta livre em Java para detecção de interações entre atributos utilizando o método da <i>multifactor dimensionality reduction</i> (MDR).
LingPipe	API em Java	Possui ferramentas de processamento de texto, que são usadas através de linguísticas computacionais	Ferramenta de mineração livre voltada para análise lingüística.
KNIME	Ferramenta livre	Baseada no eclipse	Plataforma de mineração de dados aberta que implementa o paradigma de <i>pipelining</i> de dados.
Linguagem R	Software Livre	Por ser uma linguagem funcional, o R facilita a manipulação de dados	Desenvolvida para a estatística computacional é baseada na linguagem C e fornece diversas bibliotecas para os mais variados processamentos.
Genamics Expression	Genamics Developer	Análise de proteínas de sequências de DNA	Algoritmo de análise de sequências.

Fonte: Adaptado pela autora de (Camilo e Silva, 2009)

Neste trabalho, os dados analisados a partir da plataforma computacional para Mineração de Dados, foram submetidos à linguagem R para análise estatística e visualização gráfica usando as diversas bibliotecas disponibilizadas que apresentam um bom desempenho.

Basicamente, os principais parâmetros a serem considerados na escolha de uma técnica de mineração de dados são: Que tipo de problema, de descoberta, de conhecimento, a ser solucionado: este parâmetro é obtido com a definição da tarefa de mineração de dados que deve estar de acordo com os objetivos definidos para a descoberta de conhecimento em questão. Em seguida devem-se atentar as características dos dados: a adequação da técnica de mineração de dados às características dos dados visa, principalmente, minimizar as dificuldades geralmente encontradas na transformação de dados. Na etapa de aplicação a mineração de dados pode ser aplicada como um processo de verificação onde o usuário tenta provar uma hipótese acerca da relação entre os dados, ou como um processo de descoberta onde não é feita nenhuma suposição antecipada. Existem técnicas mais propícias para o processo de verificação (análises estatística e multidimensional) e outras para o processo de descoberta (regras de associação, árvores de decisão, algoritmos genéticos e redes neurais). E o último parâmetro é a disponibilidade de ferramenta de mineração de dados: um problema de descoberta de conhecimento pode ser solucionado, em determinados casos, com a aplicação de mais de um tipo de técnica de mineração de dados. Assim, pode ser escolhida uma técnica ou outra dependendo da ferramenta disponível (FAYYAD, 1996).

2.6 ANÁLISE DE ASSOCIAÇÃO

Uma dos métodos não-supervisionados mais populares da mineração de dados é o método destinado a encontrar conjuntos de itens frequentes a partir de transações registradas em banco de dados, com a extração de regras de associação entre os itens presentes nas transações, sem levar em consideração as implicações de causalidade (TAN et al., 2009, HAN, 2011, WU et al., 2007)

A tarefa associação tem como premissa básica encontrar elementos que implicam na presença de outros elementos em uma mesma transação, ou seja, encontrar relacionamentos ou padrões frequentes entre conjuntos de dados (BRUSSO, 2000). O primeiro algoritmo eficiente de regras de associação foi o algoritmo Apriori desenvolvido por Agrawal e Srikant em 1993. O algoritmo Apriori realiza buscas sucessivas em toda a base de dados, mantendo um ótimo desempenho em termos de tempo de processamento (Agrawal & Srikant, 1994). O primeiro passo deste algoritmo é a pesquisa de conjuntos de itens frequentes, primeiro o algoritmo pesquisa todos os conjuntos de itens que aparecem com um apoio superior ao limite estabelecido pelo usuário e na fase seguinte o algoritmo constrói as regras a partir de

conjuntos de itens encontrados na primeira etapa. O algoritmo calcula a confiança de cada regra e mantém apenas aqueles em que a confiança é maior que um limiar definido pelo usuário. (DIAS,2001 ; PASTA, 2011).

A maneira de escolher as regras de associação possíveis é procurando aquelas de maior cobertura e eficácia, ou seja, procurando combinações de pares atributo-valor que possuem um mínimo de cobertura exigida e o maior grau de precisão, também conhecido como grau de confiança (WITTEN, FRANK E HALL, 2011).

O problema pode ser formalizado como se segue. Seja $I = \{i_1, i_2, \dots, i_m\}$ um conjunto de atributos binários chamados itens e seja T uma base de dados de transações, onde cada t é representada por um vetor binário, com $t[k] = 1$, se t indica a compra do item I_k , e $t[k] = 0$, caso contrário. Existe uma tupla na base de dados para cada transação. Seja X um conjunto de itens em I . É dito que a transação t satisfaz X se, para todos os itens I_k em X , $T[k] = 1$ (TAN et al., 2009).

Uma regra de associação é uma implicação da forma $X \Rightarrow Y$, onde $X \subset I$, $Y \subset I$, e $X \cap Y = \emptyset$. A regra $X \Rightarrow Y$ é válida no conjunto de transações T , com o grau de confiança c , se $c\%$ das transações em T que contêm X também contêm Y . A regra $X \Rightarrow Y$ tem suporte s em T , se $s\%$ das transações em T contêm $X \cup Y$. Se as condições forem satisfeitas $c\%$ representará o fator de confiabilidade e $s\%$ o fator de suporte (AGRAWAL et al 1993).

Assim, dado um conjunto de transações, o problema de mineração por regras de associação está em gerar todas as regras que contenham o suporte e confiança iguais ou maiores do que os valores mínimos determinados pelo usuário, referenciados como suporte mínimo e confiança mínima, respectivamente (Hand et al. 2001). O suporte de uma regra $X \Rightarrow Y$, onde X e Y são conjuntos de itens, é dado pela seguinte fórmula:

$$\text{Suporte} = \frac{\text{Frequência de X e Y}}{\text{Total de T}}$$

O numerador se refere ao número de transações em que X e Y ocorrem simultaneamente e o denominador ao total de transações. A sua confiança é dada pela seguinte fórmula:

$$\text{Confiança} = \frac{\text{Frequência de X e Y}}{\text{Frequência de X}}$$

Equação 2

O numerador se refere ao número de transações em que X e Y ocorrem simultaneamente. O denominador se refere à quantidade de transações em que o item X ocorre. O Suporte pode ser descrito como a probabilidade de que uma transação qualquer satisfaça tanto X quanto Y, ao passo que a Confiança é a probabilidade de que uma transação satisfaça Y, dado que ela satisfaz X (AGRAWAL et al 1993).

Em análises de associação aplicadas na área de marketing também se utiliza a medida Lift, definida por:

$$\text{Lift} = \frac{\text{Confiança (X} \Rightarrow \text{Y)}}{\text{Suporte (Y)}}$$

Equação 3

Essa medida permite realizar análises de independência entre variáveis e não será o foco nesta aplicação.

2.7 CLUSTERIZAÇÃO OU AGRUPAMENTO

A análise de agrupamento (clusters) busca padrões em um conjunto de dados agrupando-os em clusters de forma automática. Cada cluster é composto por indivíduos parecidos, porém não há similaridade entre clusters diferentes, em outras palavras tem-se homogeneidade dentro dos clusters e heterogeneidade entre eles. Ao se fazer uma análise de clusters tem-se a esperança de encontrar o agrupamento natural dos objetos para que os clusters formados façam sentido ao pesquisador (Macedo e Matos , 2010).

Esta técnica é utilizada quando o foco não é na predição de classes, mas quando as instâncias são divididas em grupos naturais que apresentam fatores que os relacionam e ao mesmo tempo deixam outras instâncias fora do grupo de classificação. Os grupos podem ser probabilísticos, onde o conceito de que uma instância possui certa probabilidade de pertencer a um grupo é o fator responsável pela associação (Witten, Frank e Hall 2011).

O que diferencia a clusterização da classificação é a não existência de grupos pré definidos. Para criar os clusters é necessário calcular alguma medida de semelhança, estas medidas podem ser apenas distâncias ou uma correlação entre as observações ou ainda uma ponderação destas medidas. O Quadro 3 apresenta uma comparação das principais características dos algoritmos de agrupamentos. No quadro utiliza-se a seguinte legenda: k é o número de clusters, n é o número de elementos.

Quadro 3 - Comparativo de algoritmos usados na análise de agrupamentos

	Algoritmo	Complexidade	Indicações de uso	Desvantagem / tipos de clusters
Particionamento	EM	Se k e d (dimensão) são fixas, pode ser resolvido em $O(n \sum_{k=1}^d k \log n)$	Dada uma estimativa para os parâmetros, calcula probabilidades para avaliar os novos parâmetros. Não exige a definição inicial do número de clusters ou da distância entre os pontos que formam os clusters	Encontra clusters que são de formato esférico e similar em tamanho. Pode ser lento, não é prático para modelos com grande quantidade de componentes
	K-means	$O(nkt)$ $t = nr$ iterações	Em aplicações onde se deseja que a distância para o centro do cluster seja curta. Funciona bem quando os grupos são compactos e separados. É menos sensível a valores extremos. É eficiente em grandes bases	Não pode ser aplicado em variáveis categóricas. Precisa estimar o parâmetro K . Sensível a ruídos ou valores extremos, com forte influência destes nos cálculos. Clusters esféricos e similares em tamanho
	Clara	$O(k^2 + k(n - k))$	Grande quantidade de dados. Utiliza amostras do conjunto de dados para procurar os melhores k medóides dentre essas amostras	Pode não encontrar o melhor agrupamento para todo o conjunto de dados. Assim, um bom agrupamento baseado em amostragem não significa a certeza de um bom agrupamento para todos os conjuntos de dados
	PAM	$O(k(n - k)^2)$	Procura o melhor k medóide	Proibitivo para grandes

			entre os dados usados. Em aplicações onde se deseja que a distância para o centro do cluster seja curta. Apresenta menor sensibilidade a erros e outliers	bases de dados. Muito tempo de processamento. Clusters de formato esférico e similar em tamanho
	Clarans	$O(n^2)$	Escolhe as amostras do conjunto de dados dinamicamente durante a execução	A qualidade dos resultados não pode ser garantida quando N é tem um alto valor. Assume que todos os objetos estão na memória principal
	Agnes	$O(n^2 \log n)$	Usa a abordagem aglomerativa	Tende a produzir clusters errôneos
Hierarquicos	Diana	$O(n^2)$	Usa a abordagem divisiva	Tende a produzir clusters errôneos
	Cure	$O(n^2)$	Mais robusto a outliers. Trabalha bem com grandes bases de dados sem perder em qualidade dos clusters	Ajusta-se bem para clusters com formato não esféricos
	Chameleon	$O(n^2)$	Mais eficiente do que o Cure para a descoberta de cluster de formatos arbitrários	Custo alto de processamento quando usa muitas dimensões
	Birch	$O(n)$	Usa a noção de raio ou diâmetro	Não trabalha bem se os clusters não são esféricos
	Optics	$O(\log n)$	Pode descobrir clusters de formato arbitrário	Não é muito sensível aos parâmetros. Baixo desempenho. Tamanhos e formas arbitrários
Densidade	Denclue	$O(n \log n)$	Para computar a soma das influências eficientemente, uma estrutura de grade é utilizada	Requer uma escolha cuidadosa dos parâmetros de agrupamento
	Dbscan	$O(n \log n)$, pior caso, $O(n^2)$	Descobre regiões com alta densidade e descobre regiões de formatos arbitrários. Não é muito afetado por ruídos. Identifica grupos de diferentes tamanhos	É sensível aos parâmetros de entrada. Baixo desempenho. Tem problemas quando os grupos têm densidade variada e muitas dimensões
	Sting	$O(k)$ K é nr de grids	Trabalha bem com dados espaciais	Dependente dos parâmetros de entrada e pode perder grupos naturais. Como as células são retangulares, não captura com precisão a densidade das áreas circulares. O aumento de dimensionalidade provoca o aumento exponencial do número de potenciais células da grade
Grid	WaveCluster	$O(n)$	Remove automaticamente outliers. Pode manipular dados com mais de 20 variáveis	Descobre clusters de formato arbitrário

	Clique	O $(Ck + nk) k -$ mais alta dimensão	Trabalha com dados numéricos	É de difícil aplicação por que o tamanho do grid e o valor limite de densidade de corte, são usados para todas as combinações das dimensões no data set
--	--------	--------------------------------------	------------------------------	---

Fonte: Silva (2013)

2.8 ALGORITMO PAM

O algoritmo PAM (*Partitioning Around Medoids*), como os demais métodos de agrupamento por particionamento, minimiza uma função custo em relação a um determinado vetor contendo k centroides, porém neste algoritmo estes centroides são objetos denominados medóides. Os medóides são objetos representativos de cada agrupamento e contêm os padrões onde a dissimilaridade média dos itens pertencentes a um dado agrupamento é mínima (VALE, 2005).

Este algoritmo é dividido em duas fases: A fase 1 é a construção essa fase é relativa à construção dos objetos medóides e são construídos através de k seleções de objetos representativos. O primeiro medóide é o item onde a soma das dissimilaridades entre todos os itens é mínima, os medóides seguintes são selecionados de forma a minimizar a função objetivo o máximo possível, esta função é dada por:

$$F_0 = \sum_{i=1}^n d(x_i, m(x_i))$$

Equação 4

Onde n é o total de itens no conjunto de dados, x_i é o i -ésimo item do conjunto de dados, $m(x_i)$ é o medóide mais próximo ao objeto x_i e $d(x_i, m(x_i))$ é dissimilaridade entre x_i e $m(x_i)$. Os passos para encontrar os medóides são os seguintes (VALE, 2005):

1. Considere um item x_i que não tenha sido selecionado ainda;
2. Considere um item x_j não selecionado e calcule a diferença entre a sua dissimilaridade em relação ao último objeto selecionado (D_j) com a dissimilaridade do objeto x_i selecionado no passo anterior ($d(x_j, x_i)$);

3. Se a diferença for positiva. Calcule: $C_{ij} = \max(D_j - d(x_j, x_i), 0)$;
4. Calcule o total obtido por selecionar o item x_i : $Total = \sum_{j=1} C_{ij}$;
5. É selecionado o item x_i que maximize o Total.

A fase 2 é a fase de troca , nesta fase, tenta-se melhorar o conjunto de medóides trocando os itens entre eles. Dessa forma, se houver uma minimização da função objetivo então mantém-se a troca, caso contrário ela é desfeita. O resultado final é medido pela Distância Média Final:

$$DMF = \frac{1}{n} \sum_{i=1}^n d(x_i - m(x_i))$$

Equação 5

Onde n é o total de itens no conjunto de dados, x_i é o i -ésimo item do conjunto de dados, $m(x_i)$ é o medóide mais próximo ao objeto x_i e $d(x_i - m(x_i))$ é dissimilaridade entre x_i e $m(x_i)$.

2.9 GRÁFICO DA SILHUETA

O gráfico da silhueta é uma técnica que foi proposta por Rousseeuw, em 1986, para avaliar particionamentos. Cada objeto (observação) é representado por um valor $s(x_i)$ chamado de silhueta, que é baseado na comparação da “consistência” e na “separação” de cada cluster. O agrupando inteiro é exibido combinando as silhuetas em um único gráfico, permitindo assim uma avaliação da qualidade relativa dos agrupamentos e uma avaliação da configuração dos dados. A largura média da silhueta fornece uma avaliação de validação do agrupamento e poderia ser usada para selecionar um número “adequado” de agrupamentos.

Sendo A o agrupamento ao qual o objeto x_i pertence, a dissimilaridade (ou similaridade) média $a(x_i)$ do objeto x_i em relação aos outros objetos de A é dada por:

$$a(x_i) = \frac{1}{N_A - 1} \sum_{j=1} d(x_i, x_j)$$

Equação 6

Onde N_A representa o total de itens contidos no agrupamento e $d(x_i, x_j)$ é a dissimilaridade (ou similaridade) entre o item x_i e x_j .

Considere-se qualquer agrupamento C diferente de A. A dissimilaridade (ou similaridade) média do objeto xi para todos os objetos C será dada por:

$$d(x_i, C) = \frac{1}{N_c} \sum_{j \in C} d(x_i, x_j)$$

Equação 7

Onde N_c representa o total de objetos contido no cluster C e $d(x_i, x_j)$ é a similaridade entre o item x_i e x_j .

A menor distância de dissimilaridade entre xi a um agrupamento A é dada por:

$$b(x_i) = \min_{C \neq A} d(x_i, C), \quad \forall C \neq A$$

Equação 8

Considera-se como B o agrupamento C que contém a menor distância dada acima. Esse agrupamento chamado de vizinho do objeto x_i e é o segundo melhor agrupamento possível para esse objeto. O valor da silhueta do objeto x_i , é calculada por:

$$S(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))}$$

Equação 9

O valor de $s(x_i)$ está entre -1 e 1 e pode ser interpretado da seguinte forma:

- $S(x_i) \approx 1$, objeto x_i está bem classificado no cluster A;
- $S(x_i) \approx 0$, objeto x_i está entre os clusters A e B;
- $S(x_i) \approx -1$, objeto x_i mal classificado no cluster A é mais próximo do cluster B.

O gráfico do cluster A é dado pelo gráfico dos valores da silhueta de todos os objetos pertencentes ao cluster A em ordem decrescente. Quanto mais próximo de 1 melhor é a qualidade do agrupamento. Os valores da silhueta média podem ser interpretados como se segue na Tabela 1.

Tabela 1: Valores da Silhueta média

$S(x_i)$	Descrição
0,71 - 1,00	Uma estrutura forte foi encontrada
0,51 - 0,70	Uma estrutura razoável foi encontrada
0,26 - 0,50	A estrutura é fraca e pode ser superficial. É aconselhável o uso de outros métodos para esses dados.
<0,26	Nenhuma estrutura substancial foi encontrada.

Fonte: Adaptação de Vale (2006).

2.10 COEFICIENTE DE GOWER

Outra medida de similaridade foi proposta por Gower (1971). Esta medida trata-se de um coeficiente geral de similaridade que permite a combinação simultânea de variáveis categóricas e numéricas. Este coeficiente é aplicável aos mais diversos tipos de variáveis: binárias, numéricas, nominais (multiníveis) e ordinais (multiníveis). Este índice é calculado pela expressão:

$$S_{ij} = \frac{1}{p} \sum_{i=1}^p S_i$$

Para dados categóricos (qualitativos), temos:

- S_{ij} = similaridade entre o objeto i e o objeto j;
- $S_i = 1$, quando há concordância entre os dados;
- $S_i = 0$, quando não há concordância dos dados;
- p = número de variáveis.

A semelhança entre as duas amostras é a média dos valores de similaridade para as p variáveis. As variáveis quantitativas (numéricas) são tratadas de maneira diferente. Inicialmente é calculado o desvio entre os valores de duas amostras, $|y_{i1} - y_{i2}|$, este valor é então dividido pelo desvio máximo ($|\max(x_k) - \min(x_k)|$), que pode ser calculado para esta variável utilizando a amostra disponível. Assim, a proporção é de fato uma distância padrão e subtraindo-a de um encontra-se uma semelhança, este cálculo é feito pela expressão:

$$S_i = \left[1 - \frac{|y_{ik} - y_{jk}|}{|\max(x_k) - \min(x_k)|} \right]$$

Equação 11

Onde:

- y_{ik} = valor da k-ésima variável para o objeto i;
- y_{jk} = valor da k-ésima variável para o objeto j;
- $\max(x_k)$ = valor máximo da k-ésima variável;
- $\min(x_k)$ = valor mínimo da k-ésima variável.

Ao coeficiente de Gower foi introduzido um elemento w_i (Delta de Kronecker) de flexibilidade. A comparação não é feita para as variáveis em que existe uma falta de informação para uma das duas amostras. O w_i funciona como uma espécie de “função indicadora”, informando a presença ou ausência de determinada informação, que assume

valor 0 (zero) se não há nenhuma informação que pertence a variável i de uma das duas amostras, e o valor de 1 (um) quando a informação este presente para ambas as amostras (VALE, 2005). Neste caso, o coeficiente toma a forma mostrada a baixo:

$$S_{ij} = \frac{\sum_{k=1}^p W_k S_K}{\sum_{k=1}^p W_k}$$

Equação 12

Onde:

- $W_k = 1$, quando se tem os valores da k -ésima variável para ambas amostras;
- $W_k = 0$, quando não se tem os valores da k -ésima variável para quaisquer das duas amostras;
- S_{ij} = similaridade entre os objetos i e j ;
- p = número total de variáveis.

3 MATERIAIS E MÉTODOS

A metodologia adotada neste trabalho é fundamentada no estudo de diversas tecnologias, as principais são: descoberta de conhecimento em banco de dados (*Knowledge Discovery in Database – KDD*) mineração de dados (*Data Mining - MD*) e linguagem R.

Para o estudo em questão a metodologia adotada pode ser enquadrada como um estudo de caso já que trata da aplicação de técnicas de Mineração de Dados no PDD (Programa Doce Desafio), um programa específico em educação para diabéticos. Desta forma, as conclusões do estudo não podem ser generalizadas para outros programas devido às características peculiares de cada. Tratando-se de uma pesquisa descritiva quantitativa o método de estudo de caso foi escolhido por possibilitar a observação do contexto a ser pesquisado.

A base de dados a ser estuda compreende os anos de 2001 a 2012. A população de estudo são os participantes do PDD. Trata-se de um programa de educação em saúde, com ênfase em exercícios físicos orientados, com o objetivo de promover qualidade de vida e melhora nos índices metabólicos e adesão dos diabéticos. Em paralelo, também tem como objetivo a capacitação e qualidade dos atendimentos dos profissionais da saúde envolvidos no programa. Este modelo tem bases provenientes do doutorado da Prof^a Dr^a Jane Dullius (DULLIUS, 2003).

Cada um dos participantes se inscreve em um dos 4 grupos de atendimento do PDD: 2 no Centro Olímpico, 1 em Samambaia e 1 em Sobradinho, conforme horário de disponibilidade e local de atendimento. A rotina de trabalho dos grupos é composta por 2 encontros semanais regulares, sempre nos mesmos locais, turnos e com o mesmo grupo de diabéticos. Cada dia de intervenção tem uma rotina semelhante que dura aproximadamente 150 minutos. No primeiro dia que o indivíduo procura o PDD ele recebe orientações sobre o programa, é informado dos procedimentos e convidado a fazer parte do grupo. Após sua aceitação ele faz sua ficha de inscrição, é submetido à anamneses, testes e exames laboratoriais. Dados utilizados no presente estudo foram: sociodemográficos como sexo, idade, tempo de diagnóstico da DM, nível de escolaridade, local de moradia, nível socioeconômico.

Os softwares utilizados para as análises foram o Excel conforme Figura 4 e o software free R, que é um software aberto e de domínio público, com vários recursos disponíveis, desenvolvido para Windows e ótima documentação (R Development Core Team, 2006).

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Cadastro	Data nascimento	Local de	Data ingresso no DD	Idade.ingresso	Tipo DM	Data diagnóstico	Tempo DM ao ingresso	Insulinoterapia ao ingresso?	Gênero	Bairro/cidade	CEP	Escolaridade	Profissão	Situação			
1	AB1143	26/11/1943	CO	22/11/2005	61	2	2003	2	N	M	Lago Norte	71525255	P	oficial de justiça	ativo			
2	AB0740	17/07/1940	SAM	17/03/2006	66	2	1990	15	N	M	Asa Norte	70744070	SC	dentista	aposentad			
3	AD0344	16/03/1944	SOB	25/08/2009	65	2	2008	0,666	N	F	Sobradinho		FI	costureira/dona de casa	ativa			
4	AD0650	08/06/1950	SAM	07/06/2005	54	1	1985	20	S	F	Asa Norte							
5	AD1161	08/11/1961	CO	16/07/2003	41	1	1992	11	S	F	Sobradinho	73000000	FI	do lar	ativa			
6	AD0948	07/09/1948	CO	12/09/2005	57	2	2002	3	N	M	Recanto das Emas	72507130	FI	motorista	ativo			
7	AD1050	31/10/1950	CO	08/08/2005	54	2	1997	8	N	M	Planaltina	73300000	MC					
8	AD0860	05/08/1960	CO	07/03/2006	45	2	2005	0,66	N	M	Asa Norte	70740763	SC	analista de sistemas	ativa			
9	AD0951	13/09/1951	CO	20/09/2006	55	2	2004	2	N	M	Asa Sul	70277030	SC	jornalista	aposentad			
10	AD0567	17/05/1967	CO	17/08/2009	42	2	2009	0,083	N	F	Asa Norte	70732060	SI	do lar	ativa			
11	AD0181	31/01/1981	CO	03/06/2005	24	1	2004	0,583	S	F	Asa Sul	70347090						
12	AD0660	06/06/1960	CO	09/04/2008	47	2	2006	2	N	F	Asa Norte	70754050	SC	bancária (apos)/psicóloga/artesã	ativa			
13	AD0660	06/06/1960	CO	09/04/2008	47	2	2006	2	N	F	Asa Norte	70754050	SC	bancária (apos)/psicóloga/artesã	ativa			

Figura 4 - Dados para Mineração em Excel
Fonte: Base de dados da pesquisa (2016)

Para cada item nos dados sociodemográficos foi utilizado um critério de agrupamento para o posterior cálculo de frequências e médias. A variável sexo foi classificada em masculino e feminino. A idade foi calculada segundo o seu valor médio e desvio padrão em anos. Quanto ao nível de escolaridade, os usuários foram classificados em nível fundamental e médio, completos e incompletos e nível superior. Para o tempo de duração do DM foram calculados a média e o desvio padrão em anos.

O estudo compreendeu as seguintes etapas: pré- processamento; processamento de estatísticas simples; extração de conhecimento da base de dados por meio de Mineração de Dados e análise dos padrões descobertos. A combinação de diversas técnicas amplia a visão e o entendimento dos problemas estudados, oferecendo um diferencial dos métodos tradicionais existentes no mercado. Nesse estudo, foram utilizados pacotes de softwares com interfaces amigáveis, de uso livre e capazes de manipular a base de dados. As ferramentas utilizadas são eficientes e geram novos dados de fácil interpretação. Várias técnicas estatísticas serão aplicadas à tarefa de Mineração de Dados com vistas à determinação de possíveis correlações entre variáveis do problema, associações e levantamento das variáveis mais significativas que descrevem o fenômeno.

A Estatística é a área da matemática que: estuda, coleta, organiza e interpreta dados numéricos, especialmente na análise de características da população de dados, por inferências a partir de amostras. As técnicas de estatística possuem muita importância dentro do processo de descoberta do conhecimento. As áreas de extração de conhecimento e de estatística estão fortemente relacionadas além de grande parte dos algoritmos de aprendizado de máquina fazerem uso de mecanismos disponíveis na estatística para realizar a descoberta de padrões, calcular aproximações, médias, taxas de erros e desvios. As duas disciplinas têm como objetivo encontrar padrões e regularidades nos dados (ELDER e PREGIBON, 1996; ROCHA, 1999).

Conforme o objetivo geral e os específicos propostos e alinhados com a metodologia empregada para a aquisição das informações a serem mineradas, num primeiro momento buscou-se compreender a necessidade dos gestores do PDD, a fim de poder oferecer uma solução para seus questionamentos. Para Dias (2001) na Mineração de Dados existe a possibilidade de não existir um problema real a ser solucionado, uma vez que a MD pode ser utilizada como um processo de descoberta, onde nem sempre é feito o levantamento das suposições a serem discutidas.

A fase de entendimento dos dados tem como objetivo a coleta, a descrição e entendimento dos dados e a análise da qualidade dos mesmos (SHARMA; OSEI-BRYSON; KASPER, 2012). Primeiramente entrou-se em contato com a responsável pelo PDD, solicitando o acesso à base de dados, para que fossem realizadas as etapas da Mineração de Dados. Por se tratar de base de dados secundária, não houve necessidade de parecer de Comitê de Ética em Pesquisa. Durante a fase de pré-processamento, foi realizada a limpeza dos dados do PDD e seleção de variáveis obedecendo ao objetivo do estudo. Minerar a partir de conjuntos de dados consistentes já não constitui um trabalho fácil, se estes ainda apresentarem problemas de consistência quando da coleta, podem não apenas atrasar ainda mais o processo ou mesmo inviabilizá-lo (GOMES et al., 2010).

Na etapa seguinte do processo, que consistia na mineração de dados propriamente dita, foram extraídas regras que permitem relacionar as diversas variáveis sob estudo. Analisou-se as características dos participantes para se obter uma idéia do grupo de pessoas que participam do programa. Foi analisada descritivamente também, características de desistentes do programa, procurando variáveis que possam influenciar na desistência das pessoas do programa. Para verificar uma relação entre a variação de glicemia após as atividades do

programa e algumas variáveis de interesse, foram utilizadas análises de correlação e análises de associação.

A finalidade da mineração de dados é encontrar regras de associação entre as variáveis, assim como a busca de padrões e a extração de conhecimento dos dados analisados, ou seja, mineração de dados utilizados para descoberta de informações. Para sua realização, optou-se pela aplicação das regras de associação. Para verificar as regras e padrões de associação, foi selecionado o algoritmo Apriori que faz uma varredura no conjunto de dados, procurando por subconjuntos que tenham relacionamentos que sejam frequentes e análise de agrupamentos (Clusterização) através do algoritmo PAM (Goldschmidt, 2005). Estas técnicas foram escolhidas por serem próximas entre si nos seus objetivos e por serem de maior compreensão por parte da gestora do PDD.

O algoritmo PAM foi utilizado uma vez que ele é mais eficiente para grandes conjuntos de dados se comparado com algum outro método hierárquico de clusterização e também por ser mais robusto se comparado ao algoritmo K-means. Foi o primeiro algoritmo a utilizar o método k-medoid de particionamento. Porém, como usa o objeto mais central (medoid) ao invés da média, é menos sensível a erros. Procura pelo melhor k-medoid entre os dados selecionados. Trabalha bem para pequenos grupos de dados. Tende a descobrir agrupamentos de formato esférico e de tamanho similar (HAN et al., 2011). A medida de similaridade utilizada foi o coeficiente de Gower já que é a medida indicada para casos onde há presença tanto de variáveis quantitativas como qualitativas.

A diretriz do PDD é viabilizar a autonomia do sujeito portador de DM perante sua própria condição por meio da educação. É uma troca de experiências, não só o portador, mas os profissionais de saúde também aprendem e vivenciam.

4 RESULTADOS E DISCUSSÃO

4.1 PERFIL SOCIODEMOGRÁFICO DOS PARTICIPANTES DO PROGRAMA

Nesta pesquisa, a aplicação das técnicas de mineração de dados foi apresentada através do banco de dados de um programa de educação em saúde que possuía vasto volume. Vale ressaltar que para preparação dos dados foi realizada a primeira etapa da mineração de dados que consiste na seleção de variáveis obedecendo ao objetivo do estudo. Essa etapa é abordada por diversos autores, para Steiner (2006) e Quoniam et. al. (2001), essa fase consiste na preparação dos dados, no tratamento de ruídos, dados estranhos ou inconsistentes.

Primeiramente são apresentados os resultados descritivos dos 784 participantes, iniciando-se pelas características sociodemográficas. Em seguida, destaca-se comparação entre as variáveis de interesse. Conforme Tabela 2 a base de cadastro é composta por 784 participantes sendo que 664 são do **Centro Olímpico**, 48 de **Samambaia** e 56 de **Sobradinho**, 16 ficaram em branco.

Tabela 2: Quantidade de Participantes por local de atendimento

Local	Frequência
Em Branco	16
CO	664
SAM	48
SOB	56
Total	784

A Tabela 3 apresenta a quantidade de participantes por gênero e local de atendimento, e a Tabela 4 apresenta as proporções em cada local de atendimento. Os locais de atendimento que tiveram maior participação no estudo foram o Centro Olímpico e Sobradinho, seguido por

Samambaia. Observou-se participação de 63% de mulheres nesse programa e isso acontece em todas as áreas de tratamento (CO, Samambaia e Sobradinho). A participação mais expressiva das mulheres em pesquisas sobre o tema Diabetes *Mellitus* é compatível com a maioria dos estudos nacionais e internacionais.

Nos estudos de Leite et. al. (2001) desenvolvido na cidade de Curitiba e Carolino et.al., (2008) em Maringá, apresentam proporções ainda maiores em relação a essa pesquisa, atingindo 83,30% e 84,80%, respectivamente. Na América Latina, no estudo do DEAL (Diabetes En América Latina) as mulheres representam 53,60% da amostra (STEWART et al., 2007). Já em pesquisa de abrangência nacional tiveram participação de 56,40% a 61,00% de mulheres (BRASIL, 2004c; GOMES *et al.*, 2006). Em contra partida, poucas pesquisas relatam maior participação de homens, sendo estas realizadas em países como Espanha (CAMPOS, 2009) e Itália (MUSACCHIO *et al.*, 2011). Esses dados devem ser considerados para auxiliar no planejamento de medidas para inserção do homem nos programas de Saúde, pois ao retardar a procura pelos serviços de saúde eles estarão sujeitos a sofrer complicações pela falta de tratamento e controle do DM.

Tabela 3: Quantidade de participantes por local de atendimento e gênero.

	Em Branco	CO	SAM	SOB	Total
Feminino	11	407	34	44	496
Masculino	5	257	14	12	288
Total	16	664	48	56	784

Tabela 4: Proporção de participantes por local de atendimento e gênero

	CO	SAM	SOB	Total
Feminino	61%	71%	79%	63%
Masculino	39%	29%	21%	37%
Total	100%	100%	100%	100%

A idade média dos participantes do programa é apresentada na Tabela 5 e Figura 5 abaixo.

Tabela 5: Idade média por local de atendimento

Local	Média de Idade
EmBranco	59,71
CO	57,32
SAM	60,08
SOB	64,46
Total	58,02

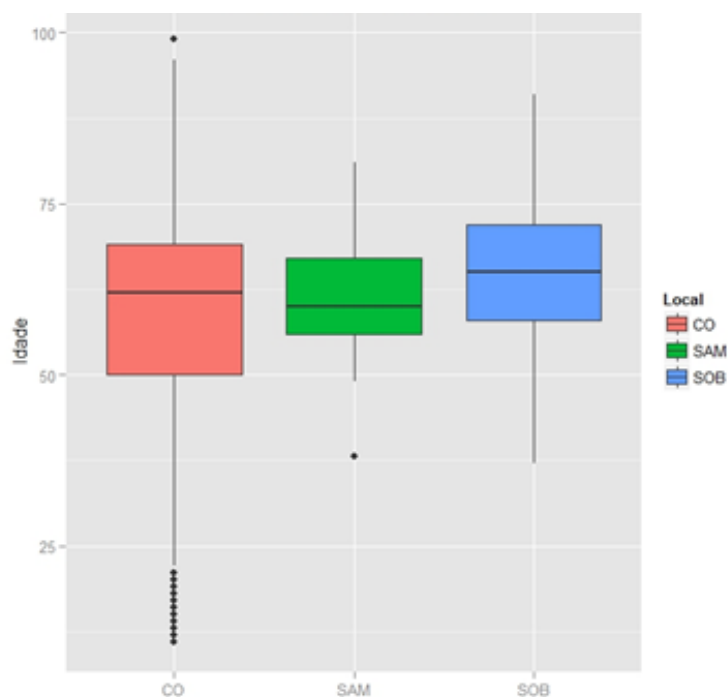


Figura 5: Idade por local de atendimento

Os participantes são, em sua maioria, pessoas idosa sendo a idade média de 58 anos. Pelo box-plot percebemos que o local de atendimento com a maior variabilidade de idade é o CO – Centro Olímpico, nele temos que 75% dos participantes tem mais de 50 anos de idade. Em outros estudos semelhantes à média de idade encontrada foi de 63,4 anos, 63,7 anos e acima de 40 anos (SAMPAIO et. al., 2008; TORRES et. al., 2011; FERREIRA E FERREIRA 2009).

A idade é uma variável importante a ser considerada em um programa educativo, pois as pessoas idosas possuem limitações de mobilidade, dificuldade no entendimento das informações e níveis de dificuldade cognitiva diferenciados para o cumprimento do plano terapêutico (BAQUEDANO et al., 2010). Para Schmidt et al.(2009), O Diabetes *Mellitus* tipo

2 é uma doença que atinge pessoas com mais idade. No entanto, pode surgir em qualquer fase, sendo mais comum o diagnóstico após os 40 anos (SBD, 2009).

Nos gráficos e tabelas apresentados anteriormente foram considerados todos que já participaram do PDD mesmo que tenham desistido. Foram excluídas do gráfico a idade de 2 mulheres que declararam a data de nascimento errada.

A idade média de ingresso dos participantes é 52 anos sendo o CO o local com a menor média de idade de ingresso conforme Tabela 6 e Figura 6

Tabela 6: Idade média de ingresso por local de atendimento

Local	Média de Idade de Ingresso
Em Branco	56,85
CO	50,87
SAM	57,34
SOB	61,75
Total	52,09

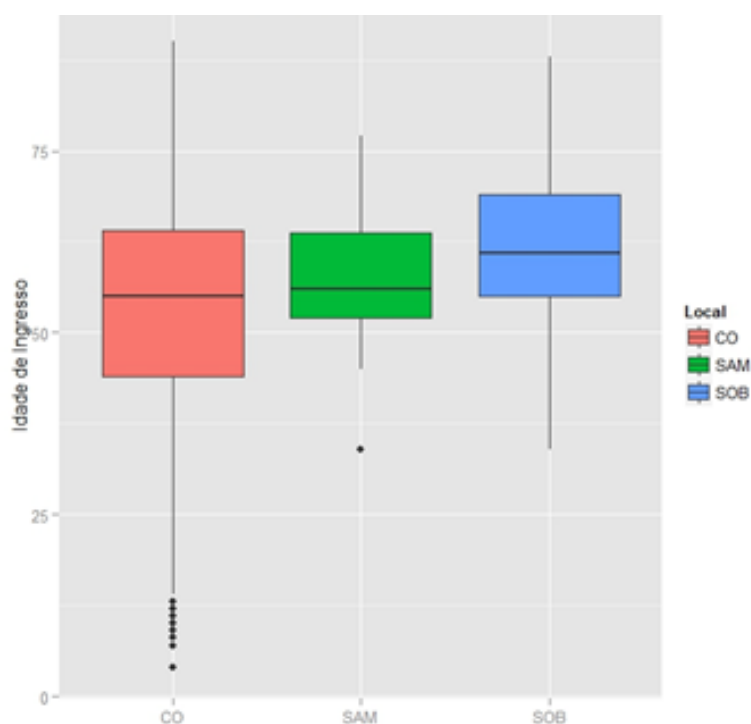


Figura 6: Idade de ingresso por local de atendimento

O tempo do diagnóstico até o ingresso ao programa é apresentado na Tabela 7 e na Figura 7 abaixo.

Tabela 7: Tempo do Diagnóstico ao Ingresso

Local	Tempo do Diagnóstico ao Ingresso (anos)
Em Branco	7,25
CO	7,22
SAM	6,76
SOB	7,73
Total	7,23

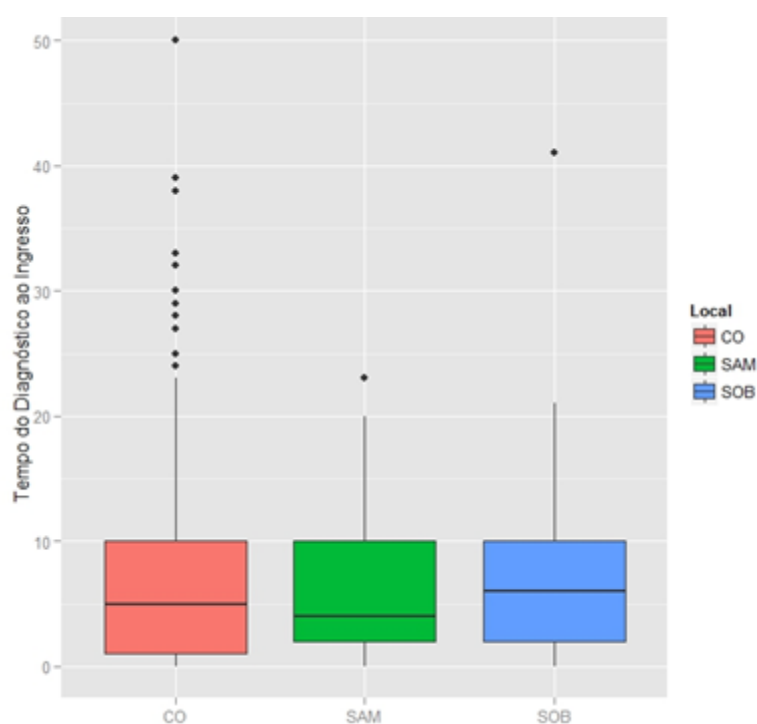


Figura 7: Tempo de diagnóstico ao ingresso por local de atendimento.

Essa variável apresenta uma informação muito interessante, pois mostra que em média os participantes levaram 7 anos para buscar o programa desde que descobriram a doença. Corroborando com esse fato, Bezie et al., (2006), afirma que a adesão de pessoas com Diabetes *Mellitus* tipo 2 é maior nas pessoas que são diagnosticadas há mais de cinco anos diferentemente daquelas com até cinco anos de diagnóstico. Esse fenômeno pode ser explicado pelo fato que com o passar do tempo, a pessoa com DM aumente seu conhecimento sobre a doença, através do contato com a comunidade e com os profissionais da saúde que os auxiliam no controle da doença e cuidados necessários no dia a dia. Ratificando esse

argumento, estudos desenvolvidos por Francioni e Silva (2007) e Pontieri e Bachion (2010), observaram-se que na fase inicial é muito comum as pessoas com DM apresentarem contrariedade e sentimentos que influenciam negativamente à adesão ao tratamento.

É importante verificar quais os tipos de Diabetes que são atendidos em geral pelo programa. As Tabela 8, Tabela 9 e Figura 8 apresentam as quantidades observadas no estudo e as proporções por local de atendimento.

Tabela 8: Quantidade de pacientes por tipo de Diabetes

Tipo de DM	Frequência
Em Branco	27
1	139
2	578
3	5
pré	35

Tabela 9: Porcentagem do tipo de diabetes por local de atendimento

	CO	SAM	SOB	Total
1	21%	2%	4%	19%
2	74%	98%	91%	76%
3	1%	0%	0%	1%
pré	5%	0%	5%	5%
Total	100%	100%	100%	100%

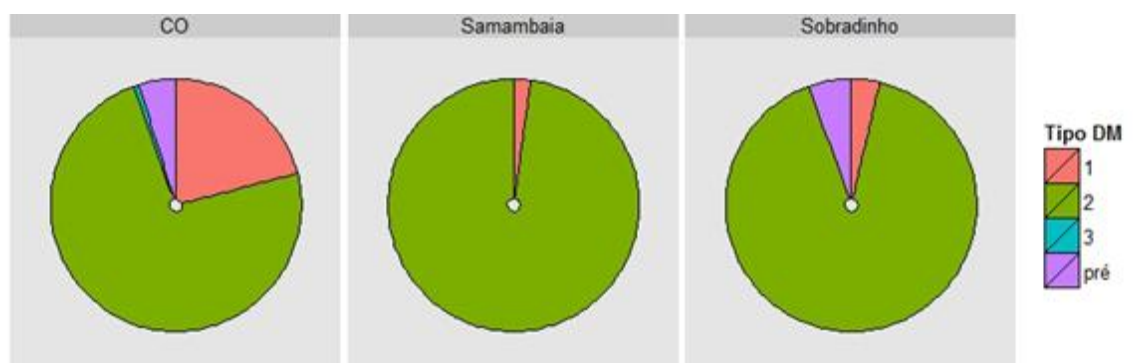


Figura 8: Porcentagem dos tipos de DM por local

Na Tabela 9 a porcentagem de pessoas com DM1 é de 19% nesse caso pessoas com DM1 não há produção de insulina pelo pâncreas tornando-se um insulino dependente, mais comum em crianças. Já no tipo 2, o pâncreas ainda produz o hormônio, mas não em quantidades suficientes para metabolizar toda a glicemia ingerida pelo organismo: atinge principalmente os adultos. A classificação para o DM3 foi utilizada pelos gestores do PDD para indicar pessoas que tiveram diabetes decorrentes de outra doença, ou acidente, ou efeito diabetogênico, etc. Por fim o pré-diabetes com 5% é a situação clínica que precede o diagnóstico do diabetes tipo 2. A grande maioria dos participantes apresentam diabetes do tipo 2 sendo que praticamente apenas no CO se encontram pacientes com outros tipos de diabetes.

Uma variável que ajuda a identificar a renda do indivíduo é a utilização de consultas particulares de saúde, pois pessoas de renda mais baixa dificilmente possuirão condições de terem consultas particulares. Não é possível generalizar, pois pessoas de alta renda também podem não utilizar consultas particulares.

A Tabela 10 apresenta a quantidade de pessoas por tipo de plano de saúde nos diferentes locais de atendimento e a na Tabela 11 a proporção do tipo de consulta por local de atendimento.

Tabela 10: Quantidade de tipo de consulta por local de atendimento

Local	Em Branco	CP	PU	Total
CO	60	486	118	664
SAM	3	8	37	48
SOB	11	15	30	56
Total	85	514	185	784

Tabela 11: Proporção do tipo de consulta por local de atendimento

Local	CP	PU	Total
CO	486	118	604
SAM	8	37	45
SOB	15	30	45
Total	514	185	699

Podemos observar também que o CO apresenta a maior quantidade de pessoas que possuem condições para utilizar consultas privadas, sendo a proporção nos outros locais de

atendimento inferior ao Centro Olímpico. Era de se esperar, pois o CO se encontra próximo de pessoas com alto poder aquisitivo (Plano Piloto). Esta característica pode ser interpretada sob dois cenários. A primeira refere-se à constatação que as pessoas que predominantemente utilizam os serviços de saúde pública, cobertos pelo serviço público de saúde são as que possuem menor renda (FREITAS et al., 2011).

A segunda poderia estar associada à maior prevalência de Diabetes *Mellitus* tipo 2 em pessoas com menor renda. A literatura traz conclusões controversas sobre esta associação. Para Moraes et al. (2010), há associação da doença e renda mensal ao contrário de Bosi et al. (2009) e Lyra et al. (2010) que não encontraram associação significativa entre estas duas variáveis, o que privilegia o primeiro cenário.

Por final, é importante verificar a escolaridade dos participantes para adequar a abordagem utilizada na divulgação do programa (Métodos de conscientização e etc.). A escolaridade foi categorizada em Fundamental Incompleto (FI), Fundamental Completo (FC), Médio Completo (MC) e Superior Completo (SC).

Tabela 12: Quantidade de escolaridade por local de atendimento

Local	Em Branco	FI	FC	MC	SC	Total
Em Branco	11	2	0	2	1	16
CO	133	78	39	156	258	664
SAM	4	32	4	6	2	48
SOB	13	23	6	10	4	56
Total	161	135	49	174	265	784

Tabela 13: Proporção de escolaridade por local de atendimento

Local	FI	FC	MC	SC	Total
Em Branco	40%	0%	40%	20%	100%
CO	15%	7%	29%	49%	100%
SAM	73%	9%	14%	5%	100%
SOB	53%	14%	23%	9%	100%
Total	22%	8%	28%	43%	100%

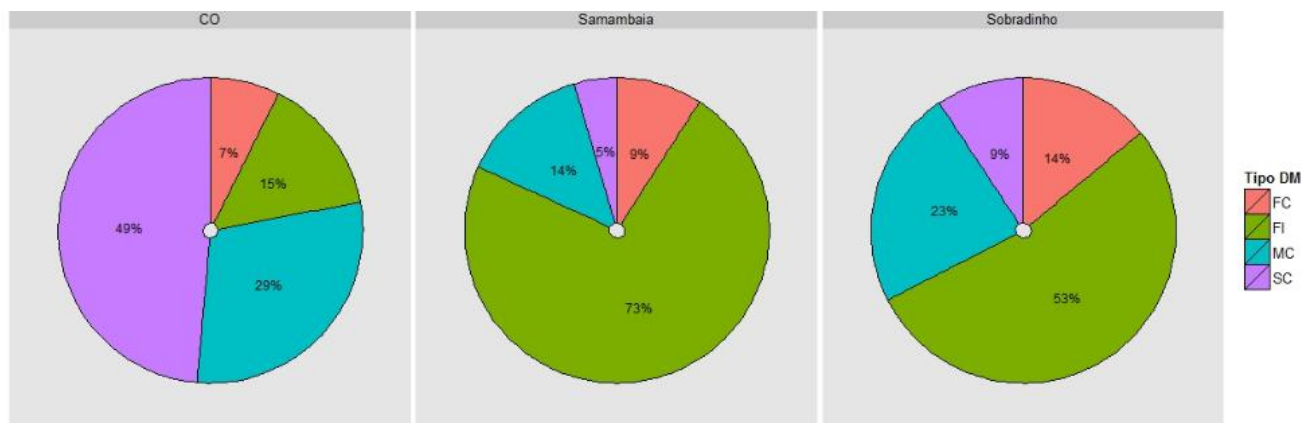


Figura 9: Porcentagem da escolaridade por local.

Podemos observar nas Tabela 12, Tabela 13 e Figura 9 o CO é o local que apresenta a maior quantidade de participantes com nível superior completo, sendo essa quantidade muito baixa nos outros locais de atendimento. Os outros locais de atendimento possuem em maioria pessoas com fundamental incompleto. Bem semelhante a outros estudos, a maior parte dos usuários com DM apresentou baixo grau de instrução (COSSON et al., 2005; COMIOTTO E MARTINS, 2006). A baixa escolaridade pode prejudicar a aprendizagem, pois, à medida que aumenta a complexidade da terapêutica medicamentosa para a doença, o paciente necessita de habilidades cognitivas mais complexas para aderir a ele (GIMENES; ZANETTI; HAAS, 2009).

O DM acomete as pessoas independente das características sociodemográficas, no entanto, Mendes (2011) defende que a educação afeta diretamente a saúde, de tal forma que baixos níveis de educação atingem negativamente a percepção do estado de saúde e ao acesso aos serviços de saúde. Isso pode ocorrer porque usuários com baixo nível de instrução apresentam maiores dificuldades para ler e entender as orientações prescritas dificultando o acesso as informações e aprendizagem quanto aos cuidados básicos que devem ser realizados diariamente para garantir o controle glicêmico (BAQUEDANO et al., 2010).

Vários estudos comprovam que há maior prevalência do Diabetes mellitus tipo 2 entre as pessoas com menor escolaridade (BOSI et al., 2009; MORAES et al., 2010). Um estudo realizado na Costa Rica, em cinco das dez unidades de saúde, apontou que 69,20% dos participantes da pesquisa tinham formação primária incompleta (menos de 8 anos de estudo). No interior paulista, 59,30% dos participantes possuíam o Ensino Fundamental incompleto (OTERO; ZANETTI; TEIXEIRA, 2007). Já nos Estados Unidos da América, a proporção de

peças com mais de oito anos de estudo foi aproximadamente 50,00% (PHILIS-TSIMIKAS et al., 2011).

4.2 CARACTERÍSTICAS DOS DESISTENTES DO PROGRAMA

A análise da característica de desistentes do programa é de extrema importância para a elaboração de mudanças visando à melhora de desempenho do programa reconhecendo quais são as barreiras que interferem na sua aderência.

Quantidade de desistentes por local de atendimento descritos na Tabela 14 foi definido como qualquer indivíduo que não compareceu a pelo menos uma aula no ano de 2012 (último ano do estudo) e não retornou nos anos posteriores. Em caso de retorno continuava a contagem dos meses a partir da data do retorno. Essa estratégia foi adotada por consideramos que as faltas podem não decorrer da desistência do programa, mas sim de alguma dificuldade transitória, tais como doença ou problemas familiares que o impediria de frequentar as sessões naquele período.

Tabela 14: Quantidade de desistentes por local de atendimento

	Em Branco	CO	SAM	SOB	Total
Continuam	10	131	2	4	147
Desistentes	6	533	46	52	637
Total	16	664	48	56	784

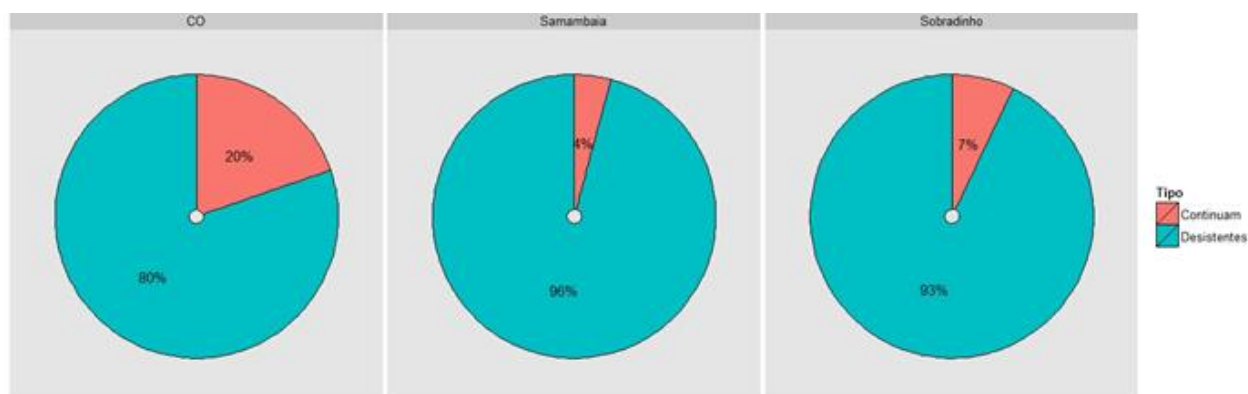


Figura 10: Porcentagem de desistentes por local

Conforme ilustrado na Figura 10 a porcentagem de desistentes é muito alta em todos os locais de atendimento, principalmente em Samambaia onde ela fica acima de 95%. A

adesão é o fator determinante para efetividade do tratamento, pois a baixa adesão pode afetar os benefícios clínicos. A adesão é uma inesgotável fonte de investigação e merece ser sistematicamente estudada, desse modo, manter indivíduos ativos fisicamente de maneira regular tem se tornado um dos grandes desafios da saúde pública (OMS, 2002).

Atestando os achados desse estudo Vancini e Lira (2004) relatam que cerca de 60% a 80% das pessoas com diabetes não seguem as prescrições mínimas para a manutenção da saúde e ressaltam a grande desistência dos indivíduos aos programas de treinamento nos primeiros meses de adesão. Dados encontrados na literatura indicam que daqueles indivíduos envolvidos em algum tipo de programa de exercício supervisionado, 50 % abandonarão em um período de seis meses e até 90% em um ano (ROBINSON e ROGERS, 1994), o que pode ser agravado com o aumento da idade, sendo ainda mais difícil a aderência de participantes adultos e idosos (DISHMAN, 1994). Estudo realizado Fecho e Malerbi (2004), apontou que apenas 19 a 30% dos pacientes portadores de DM aderem a prescrições de exercícios. Isso explica a grande desistência dos indivíduos ao programa, a idade dos participantes parece influenciar o tempo de permanência no PDD conforme Tabela 5. É importante salientar, que existe uma tendência na medida em que a idade aumenta, homens e mulheres tendem a ser menos ativos.

No entanto, a participação em programas de exercício físico supervisionado é composta por um conjunto complexo de comportamentos, o que torna difícil identificar que fatores mais contribuem para ou dificultam a aderência, particularmente entre os adultos de meia-idade e os idosos (CABRAL DE OLIVEIRA et al., 2012). Avaliando a quantidade e proporção de desistentes por sistemas de saúde obtiveram-se os seguintes resultados descritos na Tabela 15, Tabela 16 e Figura 11.

Tabela 15: Quantidade de desistentes por sistema de saúde

Local	CP	PU	Total
Continuam	87	30	117
Desistentes	427	155	582
Total	514	185	699

Tabela 16: Proporção de desistentes por sistema de saúde

Local	CP	PU	Total
Continuam	17%	16%	17%
Desistentes	83%	84%	83%
Total	100%	100%	100%

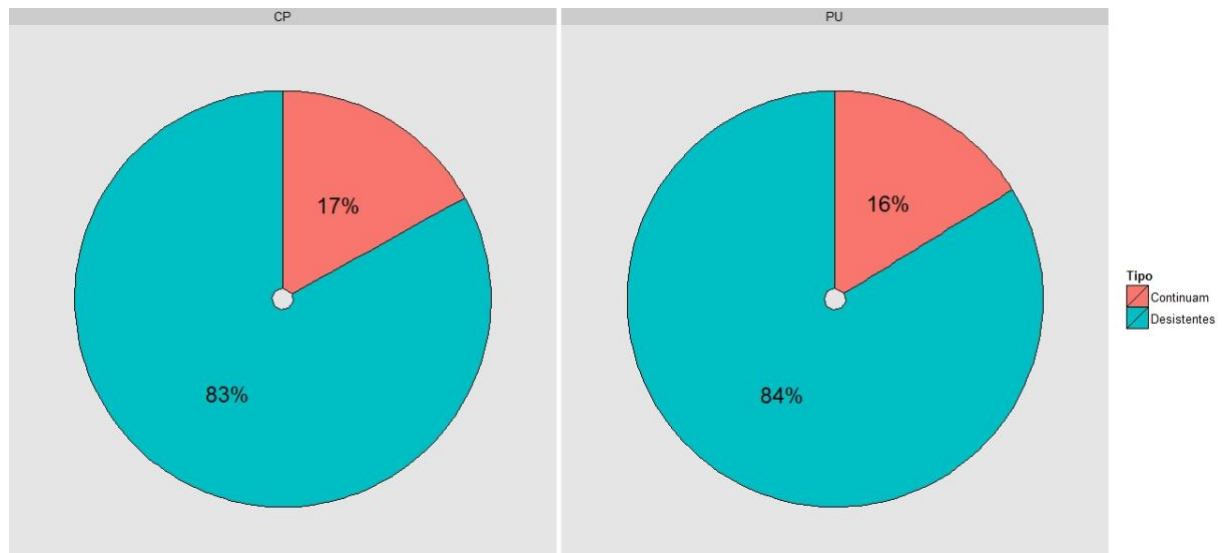


Figura 11: Porcentagem de desistentes por sistema de saúde

O sistema de saúde utilizado pela pessoa pode ajudar a diferenciar pessoas de baixa renda nos participantes. A renda é uma característica importante, pois indivíduos de baixa renda podem possuir mais dificuldade para continuar o tratamento, por questões de locomoção e etc.

Ao contrário do esperado o sistema de saúde não se mostrou influenciar a desistência ou continuação do indivíduo no programa. Uma característica que pode influenciar muito é a escolaridade desses desistentes, então é necessário avaliar a distribuição da escolaridade desses indivíduos, descritos na Tabela 17.

Local	FI	FC	MC	SC	Total
Continuam	6%	6%	21%	25%	19%
Desistentes	94%	94%	79%	75%	81%
Total	100%	100%	100%	100%	100%

Na Tabela 17 indivíduos com baixa escolaridade são os que possuem maior taxa de desistência, chegando a 94% em pessoas com Fundamental Completo e Fundamental Incompleto. As pessoas com nível superior são as que possuem menor taxa de desistência, quanto maior a escolaridade, diretamente proporcional será maior a renda e estas duas variáveis acarretarão em menor desistência, quando apresentadas dessa forma.

Por fim, é interessante verificar a quantidade de aulas que esses indivíduos participaram antes de desistirem do programa, conforme Tabela 18.

Tabela 18: Média de presenças por desistência e local de atendimento

	Em Branco	CO	SAM	SOB	Total
Continuam	19,9	49,25	33,5	16,75	46,15
Desistentes	2,83	19,35	19,08	16,96	18,98
Total	13,5	25,25	19,68	16,94	24,08

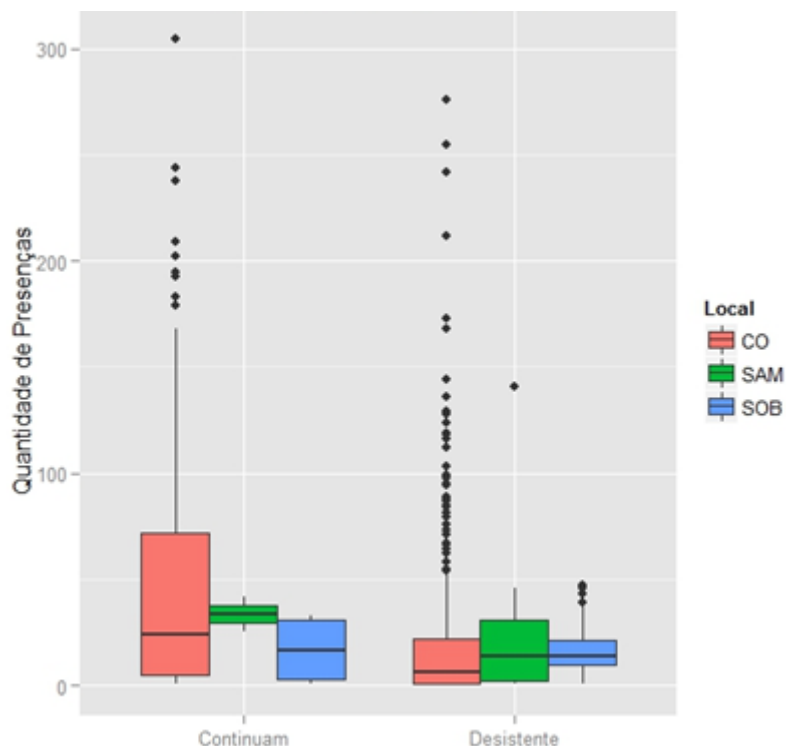


Figura 12: Presenças por desistência

Pelo box-plot Figura 12 percebe-se que os desistentes, em sua maioria, saem nas primeiras aulas, sendo no CO e Samambaia o 1º quartil equivalente a 2 aulas, isto é, 25% dos participantes desistentes só participaram de no máximo 2 aulas. Isso indica que uma grande importância nas primeiras aulas do programa.

Pelas análises descritivas as características identificadas dos desistentes são, principalmente, baixo nível de escolaridade e local de atendimento. As primeiras aulas do programa se mostraram muito importantes para incentivar a continuação do indivíduo, pois mais de 25% dos pacientes desistem nas primeiras aulas.

As variáveis Gênero e Tempo até ingresso foram analisados, mas não apresentaram resultados relevantes com proporções semelhantes em ambos os sexos e média iguais. Variáveis como Idade não foram levadas em consideração, pois a faixa etária é a mesma para os pacientes (maioria 3ª idade).

Em resumo, observou que a população é composta por pessoas idosas e com baixo nível de escolaridade, o que faz a tarefa de autocuidado mais difícil, e, portanto, o trabalho de atenção primária nos programas de educação é fundamental para evitar maiores complicações no tratamento do diabetes. Pois, usuários com melhores escores de conhecimento apresentaram maior aceitação da doença, maior adesão às práticas de autocuidado e melhoria

no controle metabólico (TORRES et al., 2009; MAULDON et al., 2006; CAMPOS et al., 2009). Portanto, ao se planejar e realizar práticas educativas, as variáveis sociodemográficas precisam ser consideradas, a fim de que as ações tenham efetividade. Políticas para a melhora do programa devem ser baseadas principalmente em conscientização de pessoas com baixo nível de escolaridade na importância do tratamento de diabetes e nas primeiras aulas do programa para uma melhor adaptação do indivíduo à nova rotina, trazendo melhores resultados no controle da doença.

4.3 INFLUÊNCIA DO TRATAMENTO NA GLICEMIA E PRESSÃO

Devido ao delineamento do estudo ser desbalanceado (muita diferença entre a quantidade de aulas de cada indivíduo, diferença entre os tempos entre aulas e diferentes tempos de entrada ao estudo), optou-se por utilizar a média das variações medidas de cada indivíduo como observação na análise.

A variação de medida da pressão arterial é dada pelo valor inicial antes da aula subtraído pelo valor medido após a aula, tendo essa medida de variação observada para 78 participantes do programa.

Tanto a pressão Diastólica quanto a pressão Sistólica são medidas nos dados de Pressão, sendo apresentadas na Tabela 19 as estatísticas obtidas:

Tabela 19: Estatísticas das variáveis Pressão Diastólica e Sistólica

	Varição Média da Pressão Diastólica	Varição Média da Pressão Sistólica
Média	-0,16	1,41
Desvio Padrão	3,58	7,42
Tamanho	78	78

O tamanho da amostra é apresentado também, pois em situação em que existem valores faltantes, eles são retirados da análise.

Realizando um teste de hipóteses para a hipótese de que a Variação média da pressão Diastólica e Sistólica é zero, obtemos na Tabela 20 o seguinte resultado:

Tabela 20: Teste de hipóteses para as variáveis de interesse.

	Varição Média da Pressão Diastólica	Varição Média da Pressão Sistólica
p-valor	0,68	0,09

Com um nível de significância de 5% ($\alpha = 0,05$) não rejeitamos a hipótese de que a variação média das pressões sejam zero, isto é, não existe evidência estatística de que os valores médios de pressão após os exercícios diminuam ou aumentem.

A hipertensão arterial é um dos principais fatores de risco para a instalação e aumento das complicações crônicas do Diabetes *Mellitus* tipo 2 e o exercício físico tem influência positiva na queda pressão arterial (MATHIEU et. al., 2008). Nessa abordagem, os valores da pressão arterial dos atendidos no PDD não estavam distante dos níveis de referência (PAS = 120-129 mmHG e PAD = 80-84 mmHg) (YAMADA et. al., 2011). Portanto, essa parece ser a razão das análises estatísticas não terem encontrado uma redução significativa nesses valores.

Utilizando a mesma metodologia utilizada para a pressão, na Tabela 21 calculou-se as estatísticas para a variação de glicemia após as aulas.

Tabela 21: Estatísticas da variável Glicemia

	Varição Média da Glicemia
Média	35,72
Desvio Padrão	36,38
Tamanho	78

A variação média após as aulas foi elevada indicando a possibilidade de o tratamento influenciar a variação da glicemia nos participantes das aulas. De forma semelhante, outros trabalhos encontraram a redução da glicemia de jejum (NEGRI et al., 2010; DIEDRICH et al., 2010). Lyra e Oliveira (2006) salientam a importância do exercício físico diário, facilitando o controle do diabetes. Deste modo, sugere-se que quanto melhor o controle glicêmico, maior a adesão ao tratamento e melhor a percepção subjetiva da QV do paciente.

Tabela 22: Teste de hipóteses para as variáveis de interesse

	Varição Média da Glicemia
<i>p</i> -valor	< 0,01

Com um nível de significância de 5% ($\alpha = 0,05$) rejeitamos a hipótese de que a variação média da Glicemia é zero, isto é, existe evidência estatística de que o valor médio da variação de Glicemia após os exercícios é afetado pelos exercícios praticados na aula. Outro programa educativo conduzido por cinco anos com 101 usuários com DM promoveu melhorias significativas sobre o controle metabólico, com a redução da glicemia, do peso e da pressão arterial (SELLI et al., 2005)

Pelas análises observou-se que os exercícios praticados não aparentam influenciar a medida de pressão após o término das aulas, porém, a glicemia se mostrou muito influenciada pelos exercícios praticados.

4.4 ANÁLISE DA INFLUÊNCIA NA VARIAÇÃO DE GLICEMIA

Para verificar a influência das variáveis de interesse na Variação de Glicemia foi necessário juntar os bancos de dados de cadastro e das medições de Glicemia, porém, o banco só possuía 48 indivíduos em comum, por isso a amostra utilizada é menor.

As variáveis que se tem interesse em verificar a influência são: ***Idade, Gênero, Frequência nas aulas, Tempo de Diabetes e Tipo de Diabetes.***

Foi utilizada a variação média da Glicemia dos participantes pelos mesmos motivos utilizados nas análises anteriores.

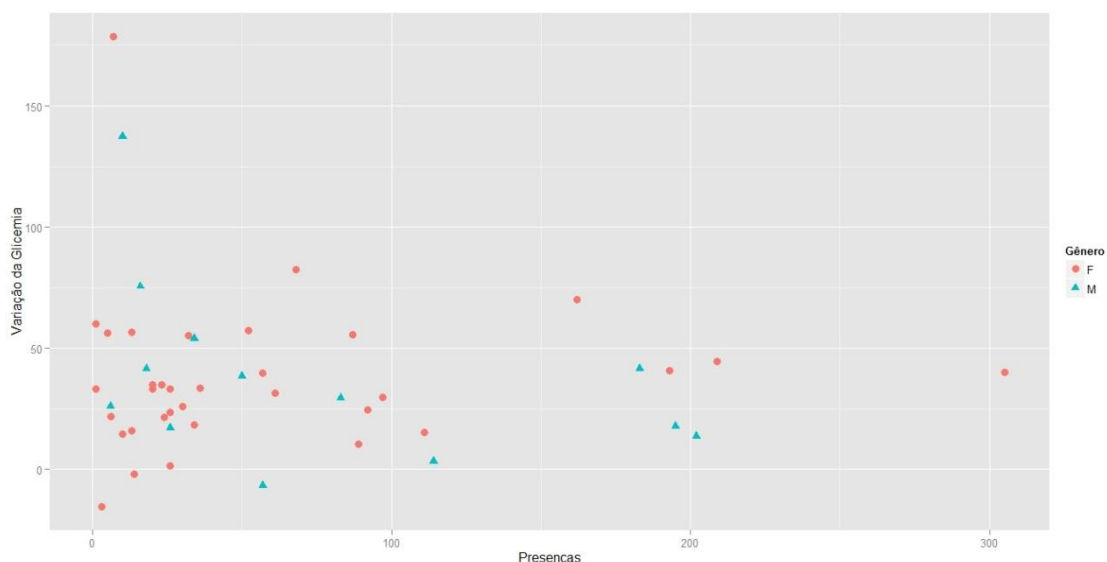


Figura 13: Diagrama de Dispersão da Variação de Glicemia pela quantidade de Presenças

Pela Figura 13 as variáveis Presenças e Gênero não parecem influenciar muito a variável variação média da glicemia, pois a medida que o número de presenças aumenta a variação de glicemia não se altera o que indica que pode não existir relação significativa.

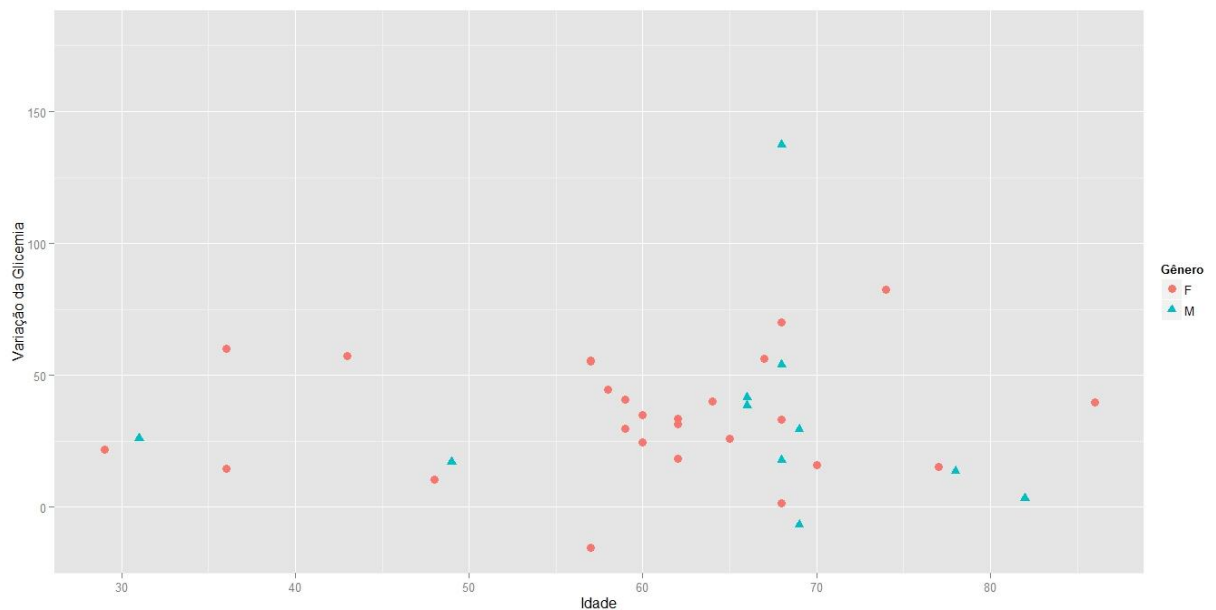


Figura 14: Diagrama de Dispersão da Variação de Glicemia pela Idade

Pela

Figura 14 a variável Idade também parece não influenciar muito a variável variação média da glicemia, pois à medida que a idade aumenta a variação de glicemia não se altera o que indica que pode não existir relação significativa.

Uma análise de regressão foi realizada com o intuito de verificar o impacto das variáveis de interesse na variação média de glicemia dos participantes do estudo. O resultado da análise é apresentado na Tabela 23:

Tabela 23: Análise de regressão das variáveis de interesse na variação média de glicemia.

Regressão	R ²	Valor F	p-valor
Modelo	0,35	1,723	0,1699

Com um nível de significância de 5% ($\alpha = 0,05$), não houve regressão, isto é, nenhuma das variáveis utilizadas apresentou alguma influência estatisticamente relevante à variação média de glicemia dos participantes do estudo. O R^2 é uma medida que varia de 0 a 1, seu valor 0,35 é outra evidência de que a variabilidade explicada pela regressão foi muito baixa. Devido à grande quantidade de valores faltantes só 26 observações dos 48 foram utilizadas na regressão final.

A utilização de um modelo de regressão linear pareceu a estatística adequada para o objetivo proposto, visto tratar-se de um modelo estatístico que pode ser usado para prever o comportamento de uma variável dependente (quantitativa), a partir de um conjunto de variáveis independentes que se mostram relevantes para o estudo, tendo em conta ainda que a regressão linear pode ser usada para modelar a relação funcional entre as variáveis, independentemente de existir ou não uma relação de causa-efeito (AMORIM, 2009).

Ainda segundo Amorim (2009), outros estudos têm evidenciado diferenças relativamente à utilização de estratégias de análise da variação glicêmica, quando se referem a variáveis sociodemográficas. Ressalta, no entanto, que diferentes instrumentos de avaliação são usados nos diferentes estudos, tornando-se, deste modo, difícil, comparar resultados da forma mais correta.

Num sentido semelhante ao presente estudo, Amorim (2009) comparou os valores médios da HgA1c, tendo em conta o gênero (teste t de Student), permitiu verificar que não existem diferenças significativas entre os resultados obtidos por homens e mulheres, tendo os homens apresentado um valor médio de 8,2% (dp=1,6) e as mulheres um valor médio de 8,0% (dp=1,6) e em relação à idade, não foi possível encontrar correlação significativa com HgA1c. Do mesmo modo, não existe evidência estatística que nos permita afirmar a existência de dependência entre as variáveis gênero e idade e a variação glicêmica nos diabéticos em estudo.

E ao contrário dos achados, Amorim (2009) observa-se uma correlação positiva e significativa, entre o tempo de doença em anos e os valores de HgA1c, observando-se, ainda, a existência de diferenças nos resultados entre os indivíduos com o diagnóstico há mais tempo, quando se distinguem os indivíduos que têm a doença há 10 anos ou menos e os que a têm há mais de 10 anos, com os indivíduos que têm o diagnóstico há menos tempo a apresentar um valor inferior em relação aos que o têm há mais tempo.

Portanto, não é possível, neste estudo, estabelecer relações entre a variação glicêmica e a maior parte das variáveis sócio demográficas.

4.5 APLICAÇÃO DAS REGRAS DE ASSOCIAÇÃO

As regras de associação é uma técnica descritiva utilizadas para descobrir as relações entre as variáveis qualitativas de um banco de dados (MCNICHOLAS; MURPHY; OREGAN, 2008). A relevância das regras é mensurada através das medidas de lift e confiança. Como ela é uma técnica utilizada em variáveis categóricas, às variáveis numéricas de interesse foram categorizadas, sendo elas:

Idade: [0,19),[19,30),[30,45),[45,60),[60,100);

Tempo com DM: [0,2),[2,4),[4,6),[6,8),[8,10),[10,20);

Presenças: [0,10),[10,25),[25,50),[50,100),[100,150),[150,200), [200,300);

Na técnica de Regras de Associação é necessário definir um grau de suporte e confiança mínimo que deseja encontrar, afim de que não se encontrem relações espúrias nem fracas. Neste trabalho utilizamos como mínimo de Suporte 50% e como mínimo de confiança 50%, visto que na Figura 15 as maiores partes das associações possuem pouca confiança ou pouco suporte.

O suporte representa o percentual de transações em que todos os itens contidos na regra estão presentes. Já a confiança expressa à força da regra, ou a chance de acerto da regra, indicando a probabilidade do lado direito da regra acontecer dado que o lado esquerdo da regra também ocorre, ou seja, a confiança é utilizada para “filtrar” as regras, permitindo somente as regras com confiança superior a inicialmente predefinida. E para medir a razão da confiança pelo percentual de transações cobertas pelo lado direito da regra existe o *Lift*. Ele varia de 0 a $+\infty$, quando o *lift* é maior que 1, o lado direito da regra ocorre com mais frequência nas transações em que o lado esquerdo ocorre. Agora quando o *lift* é menor que 1, o lado direito é mais freqüente nas transações em que o lado esquerdo não ocorre. Assim as regras que possuem *lift* maior que 1 são mais interessantes que as demais e maior deverá ser a relação entre os dois da regra (FERREIRA, 2014).

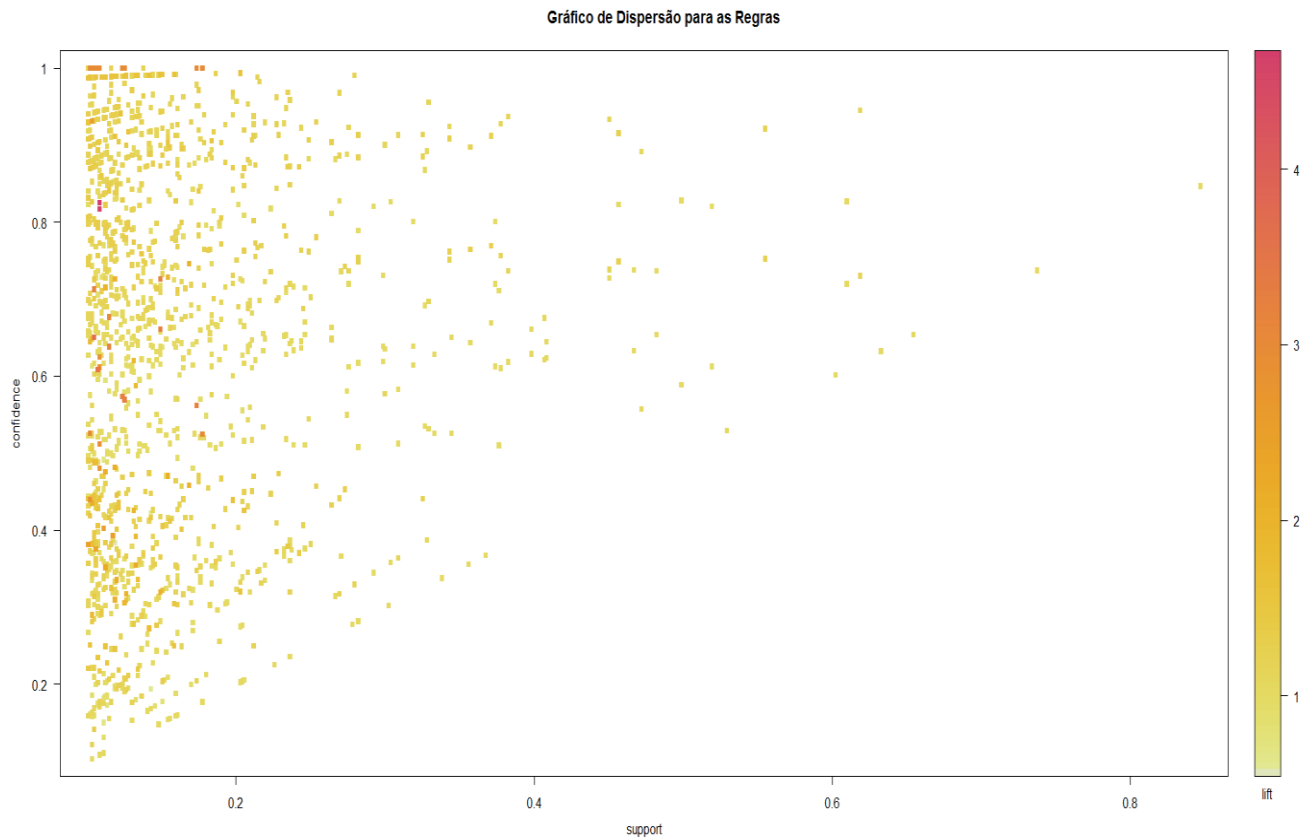


Figura 15: Confiância e Suporte das regras de Associação

Tabela 24: Associações das variáveis no Cadastro

Associação		Suporte	Confiância	Lift
Sistema de Saúde = CP	-> Local de Atendimento = CO	61,86%	94,55%	1,116
Insulinoterapia ao entrar = N	-> TipoDM = 2	55,48%	92,16%	1,250
Local de Atendimento = CO		84,69%	84,69%	
TipoDM = 2	-> Local de Atendimento = CO	60,97%	82,70%	0,976
Genero = F	-> Local de Atendimento = CO	51,91%	82,06%	0,969
TipoDM = 2	-> Insulinoterapia ao entrar = N	55,48%	75,26%	1,250
TipoDM = 2		73,72%	73,72%	
Local de Atendimento = CO	-> SisteSaude = CP	61,86%	73,04%	1,116
Local de Atendimento = CO	-> TipoDM = 2	60,97%	71,99%	0,976
SisteSaude = CP		65,43%	65,43%	
Genero = F		63,27%	63,27%	
Insulinoterapia ao entrar = N		60,20%	60,20%	0,969
Presença = (0,10]		52,93%	52,93%	

As regras de associação obtidas com confiança e suporte de 50% estão apresentadas na tabela Tabela 24. Percebe-se que existe uma confiança muito forte de que pessoas com sistema de saúde privado utilizam o Centro Olímpico para participarem do programa, mas verificando a tabela abaixo vemos que pessoas sem plano de saúde também possuem uma alta confiança de utilizar o Centro Olímpico, a diferença é que não existe um suporte alto para essa afirmação.

Tabela 25: Associações das variáveis Sistema de Saúde vs Local de Atendimento

Associação		Suporte	Confiança	Lift
Sistema de Saúde = CP	-> Local de Atendimento = CO	61,99%	94,55%	1,12
Sistema de Saúde = PU	-> Local de Atendimento = CO	15,05%	63,78%	0,75
Sistema de Saúde = PU	-> Local de Atendimento = SAM	4,72%	20,00%	3,27
Sistema de Saúde = PU	-> Local de Atendimento = SOB	3,83%	16,22%	2,27
Sistema de Saúde = CP	-> Local de Atendimento = SOB	1,91%	2,92%	0,41
Sistema de Saúde = CP	-> Local de Atendimento = SAM	1,02%	1,56%	0,25

Uma associação interessante é a de que pacientes que não utilizavam Insulinoterapia ao entrar possuem Diabetes Melitius tipo 2 com 92% de confiança. Pela Tabela 26 verifica-se que essa associação de fato é relevante, pois pessoas que utilizavam Insulinoterapia ao entrar apresentam em maior parte Tipo DM = 1.

Tabela 26: Associações das variáveis Insulinoterapia ao entrar vs Tipo de Diabetes Melitius

Associação		Suporte	Confiança	Lift
Insulinoterapiaaoentrar = N	=> TipoDM = 2	55,48%	92,16%	1,25
Insulinoterapiaaoentrar = S	=> TipoDM = 1	17,73%	52,45%	2,96
Insulinoterapiaaoentrar = S	=> TipoDM = 2	14,29%	42,26%	0,57
Insulinoterapiaaoentrar = N	=> TipoDM = pré	4,46%	7,42%	1,66

A associação de Tipo DM = 1 com Local de Atendimento = Centro Olímpico apresenta forte confiança, porém pode ser devido a poucas observações em outros locais de atendimento, visto que os outros possuem baixo suporte em suas afirmações.

Tabela 27: Associações das variáveis Tipo de Diabetes Melitius vs Local de Atendimento

Associação		Suporte	Confiança	Lift
TipoDM = 1	=> Local de Atendimento = CO	17,35%	97,84%	1,16
TipoDM = pré	=> Local de Atendimento = CO	3,95%	88,57%	1,05
TipoDM = 2	=> Local de Atendimento = CO	60,97%	82,70%	0,98
TipoDM = 2	=> Local de Atendimento = SOB	6,38%	8,65%	1,21
TipoDM = 2	=> Local de Atendimento = SAM	5,74%	7,79%	1,27

A associação de Gênero = Feminino com Local de Atendimento = Centro Olímpico também apresenta forte confiança e suporte significativo, porém pode ser devido a poucas observações em outros locais de atendimento, visto que os outros locais também possuem baixo suporte em suas afirmações.

Tabela 28: Associações das variáveis Gênero vs Local de Atendimento

Associação		Suporte	Confiança	Lift
Gênero = M	=> Local de Atendimento = CO	32,78%	89,24%	1,05
Gênero = F	=> Local de Atendimento = CO	51,91%	82,06%	0,97
Gênero = F	=> Local de Atendimento = SOB	5,61%	8,87%	1,24
Gênero = F	=> Local de Atendimento = SAM	4,34%	6,85%	1,12
Gênero = M	=> Local de Atendimento = SAM	1,79%	4,86%	0,79
Gênero = M	=> Local de Atendimento = SOB	1,53%	4,17%	0,58

As regras observadas trazem como tendência de que pessoas com sistema de saúde privado utilizam o Centro Olímpico para participarem do programa sendo que estes formam a grande parte do público ingressante do PDD. Os resultados também indicaram que um dos principais fatores para a escolha do programa é sua localização, o C.O fica no centro da cidade. Outro resultado da análise demonstra que os ingressantes deste programa são em maioria feminino, com confiança de 82,06% Tabela 28.

4.6 ANÁLISE DE AGRUPAMENTOS (CLUSTERIZAÇÃO)

A análise de agrupamento (clusters) busca padrões em um conjunto de dados agrupando-os em clusters de forma automática. Uma ferramenta importante na avaliação da qualidade de agrupamentos em sistemas de reconhecimento de padrões é conhecida por *silhouettes* (ANTONELLI et. al., 2013). A descoberta do conhecimento em base de dados é o tema de Scoss (2006) em seu trabalho de especialização, a autora faz uso da tarefa de clusterização para análise do desempenho dos docentes da Universidade do Extremo Sul Catarinense, com objetivo de analisar o perfil dos docentes no contexto da Avaliação Institucional em uma base de dados que continha 36.672 instâncias e 21 atributos. Na tarefa de clusterização foram definidos quatro clusters, sendo estes o número de áreas de 106 conhecimento disponíveis na Universidade, área de licenciatura, área de Saúde e Biológicas, área de Sociais e Aplicadas e área de Engenharia e Tecnologia.

Kampff (2009) define que a clusterização busca descobrir conhecimento de forma indireta, a partir da identificação de grupos de dados com características semelhantes. O objetivo desta técnica consiste em identificar agrupamentos de dados que podem ser classificados em uma classe comum, descoberta no processo de clusterização.

Leite (2002) ressalta que na mineração de dados, através do Data Mining, é possível se formar grupos de pessoas a partir da base de dados que se tenha, com tendências para determinadas coisas ou situações, o que abre campo para que uma empresa lance um produto ou serviço destinado a esse público.

Para realizar a identificação dos clusters inicialmente foi realizado uma avaliação das variáveis que compõem o banco de dados “Cadastro”. Neste passo buscou-se selecionar variáveis que não apresentavam um número muito grande de respostas em branco e também que não tivessem muitas categorias com frequência muito baixa ou próxima de zero, já que são dois aspectos que podem influenciar de maneira negativa os algoritmos de agrupamento e conseqüentemente afetar os resultados. Diante desses critérios, foram selecionadas as seguintes variáveis a serem trabalhadas no processo de agrupamento: Local de atendimento, Idade de ingresso, Tipo de Diabetes Mellitus (DM), Tempo DM ao ingresso, Insulinoterapia ao ingresso, Gênero, Escolaridade, Sistema de saúde e Nº de presenças.

Para realizar o processo de agrupamento trabalhou-se apenas com os indivíduos que tinham o cadastro totalmente preenchido para essas 9 variáveis, que no caso foram 530

peças. O algoritmo PAM foi utilizado uma vez que ele é mais eficiente para grandes conjuntos de dados se comparado com algum outro método hierárquico de clusterização e também por ser mais robusto se comparado ao algoritmo K-means. A medida de similaridade utilizada foi o coeficiente de Gower já que é a medida indicada para casos onde há presença tanto de variáveis quantitativas como qualitativas.

O primeiro passo foi determinar o número ideal de clusters no agrupamento dessas 9 variáveis. Como no algoritmo PAM o número de clusters deve ser definido previamente, foram ajustados agrupamentos com 2 a 10 clusters e seus valores de silhueta média foram avaliados. A Figura 16 mostra esses resultados.

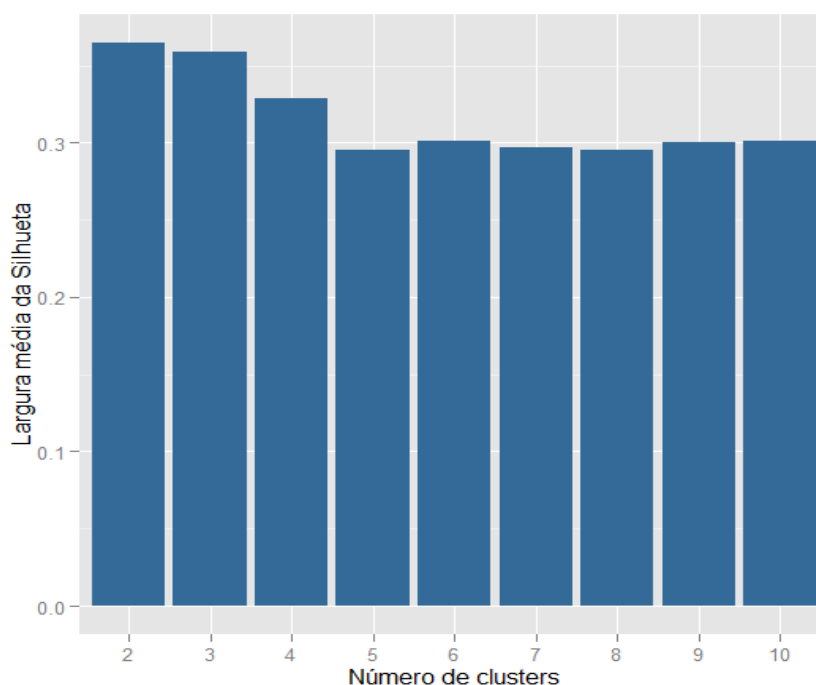


Figura 16: Resultados de silhueta média para agrupamentos com 2 a 10 clusters com o conjunto de 9 variáveis

Os valores da silhueta indicam que o agrupamento que teve melhor resultado foi com 2 clusters. Entretanto, o valor da silhueta média foi de aproximadamente 0,36, o que indica que a estrutura desses grupos é fraca e superficial, logo os resultados não são satisfatórios. Esse cenário pode ser um indicativo de que uma ou mais variáveis estejam inserindo uma indeterminação muito forte no conjunto de dados, nesse caso talvez seja melhor trabalhar com um subconjunto dessas 9 variáveis.

Para determinar qual conjunto de variáveis apresenta melhores resultados, foram simulados para cada subconjunto possível das 9 variáveis selecionadas inicialmente os resultados dos agrupamentos com 2 a 10 clusters e seus valores de silhueta média foram

avaliados. O conjunto de variáveis que apresentou melhores resultados foi: Idade de ingresso, Tipo de Diabetes Mellitus (DM), Tempo DM ao ingresso, Insulinoterapia ao ingresso e N° de presenças. A Figura 17 mostra os valores de silhueta média para os agrupamentos realizados nesse conjunto.

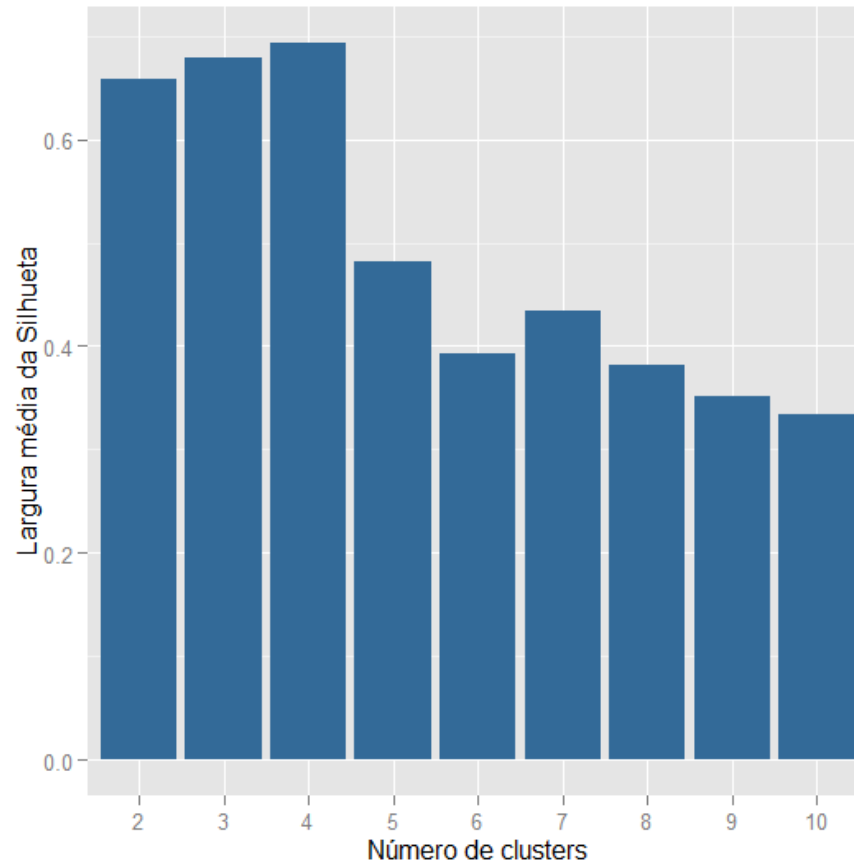
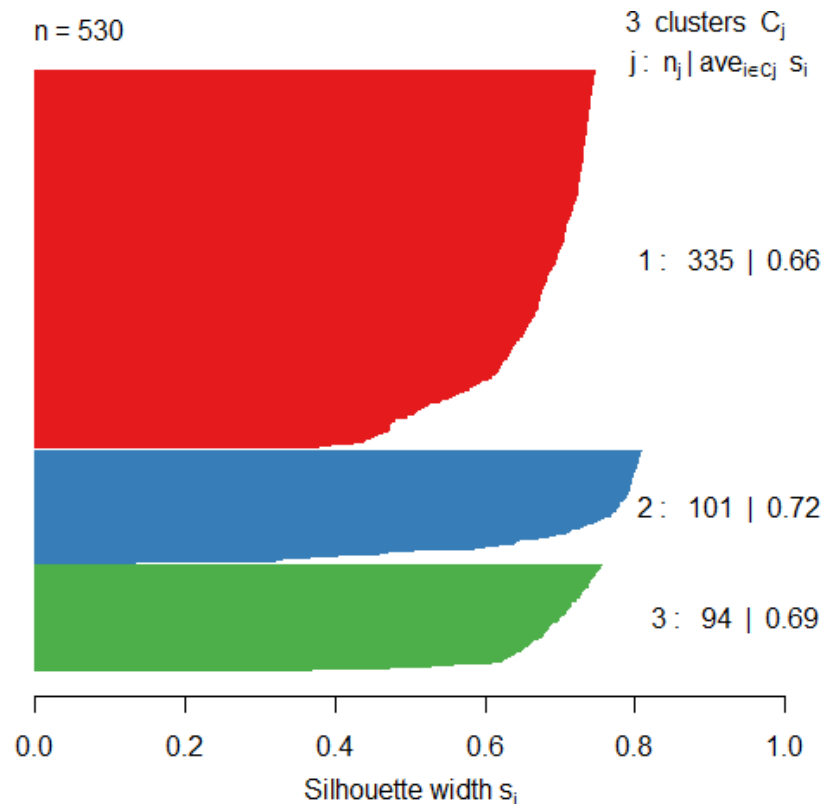


Figura 17: Resultados de silhueta média para agrupamentos com 2 a 10 clusters com o conjunto de 5 variáveis.

Avaliando os valores médios de silhueta os agrupamentos com 3 e 4 clusters foram os que tiveram melhores resultados, 0,68 e 0,69 respectivamente. Os valores são bem próximos e indicam que nos dois casos os grupos apresentam uma forte estrutura de similaridade. Como o intuito do processo de clusterização é resumir o máximo possível a informação em busca de padrões, optou-se por trabalhar com o agrupamento de 3 clusters, já que este reduz mais o conjunto de dados. A Figura 18 apresenta a silhueta para cada um dos três clusters determinados.



Average silhouette width : 0.68

Figura 18: Gráfico da silhueta para o agrupamento com 3 clusters.

O primeiro grupo é composto por 335 (**63%**) indivíduos, já o segundo por 101 (**19%**) e o terceiro por 94 (**18%**) indivíduos. O segundo grupo obteve os melhores resultados com silhueta média de 0,72. Já o primeiro e o terceiro grupos tiveram valores menores de silhueta, mas ainda assim bastante satisfatórios, 0,66 e 0,69 respectivamente. De maneira geral, o valor médio da silhueta para os três grupos, que foi igual a 0,68, mostra que a alocação dos indivíduos nesses 3 clusters é confiável.

As características para similaridade são transmitidas através dos medóides encontrados em cada grupo, ou seja, os indivíduos representativos em cada grupo, que estão apresentados na Tabela 29. Esse tipo de informação poderia ser importante na identificação do critério de agrupamento dos indivíduos, para o desenvolvimento de alguma atividade para a qual fosse interessante reunir os pacientes mais "parecidos entre si", como é o caso das práticas de educação em saúde ou das terapias em grupo.

Tabela 29: Indivíduos representativos (medóides) em cada grupo

Grupo	Idade de ingresso	Tipo de Diabetes Mellitus (DM)	Tempo DM ao ingresso	Insulinoterapia ao ingresso?	Nº de presenças
1	61	2	4	Não	16
2	19	1	6	Sim	6
3	13	2	10	Sim	13

5 CONCLUSÃO

O KDD é um processo de descoberta de conhecimento que tem como objetivo extrair conhecimento a partir de grandes bases de dados. A etapa mais importante deste processo é o Data Mining que se caracteriza pela existência de um algoritmo que diante da tarefa proposta será eficiente em extrair conhecimento explícito e útil em banco de dados, ou seja, a mineração de dados é a fase que transforma dados puros em informações úteis. Como muitas vezes esse conhecimento está oculto em uma grande base de dados, é necessário recorrer a um sistema de tecnologia da informação para auxiliar esse processo de extração de conhecimento.

Neste contexto, surgiu a motivação pela utilização de uma metodologia que possa explorar as informações dos usuários do Programa de Educação - Doce Desafio. O desenvolvimento deste estudo possibilitou analisar dados e informações dos atendimentos realizados pelo PDD, por meio da aplicação das técnicas de mineração de dados, utilizando, apresentando o perfil dos usuários do programa e dos desistentes. Nessa linha, o trabalho apresentou resultados que podem ser facilmente interpretados e ajudam no entendimento de características de nosso público alvo, contribuindo assim para a gestão e organização de campanhas dirigidas a esses perfis. Além disso, confirmou a necessidade do planejamento, avaliação e continuidade das ações educativas.

Foi considerado uma razão limitante deste estudo o fato de os dados terem sido obtidos através de uma revisão de prontuários e, devido a este motivo, nos deparamos frequentemente com informações incompletas. Foram encontradas regras válidas confirmando a validade do KDD, mas não foi encontrada regra nova válida, na opinião da especialista.

Esse estudo demonstra, que se utilizadas de forma adequada estas ferramentas podem auxiliar na tomada de decisões e elaboração de ações para a promoção de saúde e consequente melhoria da qualidade de vida da população.

A pesquisa apresentou importantes análises sobre o perfil dos usuários e desistentes do PDD, por meio da aplicação das técnicas de estatísticas e de Mineração de Dados implementadas na ferramenta R, Associação e Clusterização. Os experimentos apresentam uma importante contribuição em termos de quais aspectos são característicos para os que continuam no programa e também para os desistentes. A identificação destas características fornece subsídios para a compreensão do problema como um todo.

Uma das principais características que pode ser observada é que a maioria dos participantes são pessoas idosa sendo a idade média de 58 anos. O programa tem participação de 63% de mulheres, A idade média de ingresso dos participantes é 52 anos, em média os participantes levaram 7 anos para buscar o programa desde que descobriram a doença.

Identificou-se que a maioria dos pacientes atendidos pelo programa são Diabéticos do Tipo II e que mais de 75% dos participantes levam pelo menos 7 anos desde a descoberta da doença até a busca ao programa. Outra característica interessante é que as primeiras aulas se mostraram extremamente importantes para a continuação de um indivíduo no programa.

Das inferências realizadas, notou-se que a variação média da pressão após as aulas não se apresentou estatisticamente significativa, porém a variação média da glicemia apresentou diferença significativa após o uso do tratamento. Na análise de regressão a tabela Anova indicou a falta de relação entre as variáveis estudadas.

Apesar disso, os resultados sugerem que providências são necessárias para aperfeiçoar o programa, no que concerne aos efeitos fisiológicos perseguidos e incremento da adesão dos pacientes. É importante que os métodos de trabalho utilizados sejam repensados. Pode-se, por exemplo, aconselhar melhor treinamento dos pacientes no controle da intensidade, frequência e duração dos exercícios. Enfim, deve-se implantar um sistema para melhor acompanhar as atividades efetivamente realizadas pelos pacientes.

Os objetivos propostos foram alcançados, as associações encontradas apresentam características que devem ser mais exploradas para efeito de adequações de rotinas de trabalho ou fomento a comportamentos de autocuidado, além de outras intervenções, nos níveis primários, secundários ou terciários, na área da saúde. Os resultados deste trabalho indicam características importantíssimas dos participantes do programa de Diabéticos, ficando agora a critério de um especialista a desenvolver políticas que possam alterar essas realidades e melhorar o desempenho do programa.

Os testes mostraram útil as combinações de diversas técnicas, para descobrir padrões comportamentais que não seriam de fácil identificação por meios comuns utilizados atualmente na área da saúde. A ferramenta utilizada para a implementação dos algoritmos, sobre os dados extraídos no estudo de casos, demonstrou-se útil na execução e apresentação dos resultados para atender as exigências dos requisitos desta pesquisa.

A metodologia utilizada neste trabalho pode ser aplicada em diversos cenários dentro da área da saúde, visando a otimização dos processos e agilidade na tomada de decisões pelo bem estar da saúde, especialmente em ações preventivas. Esta seria uma boa proposta para a continuação dos esforços aqui demonstrados.

Concluindo, o presente estudo apresenta uma estruturação de avaliação de programa educativo em DM útil para conhecer seus efeitos no controle da doença e no gerenciamento do autocuidado. Além disso, permite discutir os possíveis limites e opções de aperfeiçoamento do processo de educação em saúde, associado a programas para diabetes. No entanto, ainda são necessários estudos visando identificar aspectos pedagógicos eficazes no aumento da motivação e da adesão dos portadores de DM aos programas de intervenção, compatibilizando, desta forma, conhecimento e práticas pessoais promotoras da saúde.

Como possíveis trabalhos futuros, a partir desta pesquisa, sugerem-se: Desenvolver um formulário eletrônico para coleta de informações e a armazenagem dos dados, possibilitando a geração dos arquivos no formato apropriado para a Mineração de Dados e a visualização dos resultados da mineração. Implementação de algoritmos de Mineração de Dados junto à ferramenta oportunizando ao próprio gestor elaborar sua mineração. Desenvolver uma plataforma amigável que implemente o método proposto em um microcomputador com uma configuração básica a ser utilizado nos programas de educação em saúde.

6 REFERÊNCIAS

ADA. American Diabetes Association (2010). Standards of Medical Care in Diabetes - 2010. *Diabetes Care*, 33(supl. 1), 11-61.

ADA. American Diabetes Association. Standards of medical care in diabetes-2015. *Diabetes Care* v. 38, sup 1. 2015.

AGRAWAL, R; IMIELINSKI, T; SWAMI, A. Mining Association Rules between Sets of Items in Large Databases. ACM SIGMOD CONFERENCE ON MANAGEMENT OF DATA, p. 207 – 216, Washington, DC, USA, 1993. ACM Press - New York, NY, USA.

AGRAWAL, R; SRIKANT, R. Fast Algorithms for Mining Association Rules. Bocca, J. B; Jarke, M; Zaniolo, C, editors, PROC. 20TH INT. CONF. VERY LARGE DATA BASES, VLDB, p. 487–499, Washington, DC, USA, 12–15 1994.

AMORIM, Maria Isabel Soares Parente Lajoso. Para lá dos números... Aspectos psicossociais e qualidade de vida do Indivíduo com Diabetes mellitus tipo 2 - Tese de doutorado - instituto de ciências biomédicas Abel Salazar; icbas. 2009

ANTONELLI, D. et.al., Analysis of diabetic patients through their examination history. Italia: Expert Systems with Applications, 2013.

BAQUEDANO, I. R.; SANTOS, M. A.; TEIXEIRA, C. R. S.; MARTINS, T. A.; ZANETTI, M. L. Fatores relacionados ao autocuidado de pessoas com diabetes mellitus atendidas em serviço de urgência no México. Revista da Escola de Enfermagem da USP, v. 44, n. 4, p. 1017-1023, 2010.

BARSAGLINI, R. A. As representações sociais e a experiência com o diabetes: um enfoque socioantropológico. Rio de Janeiro: Fiocruz, 2011.

BAYAM E, LIEBOWITZ J, AGRESTI W. Older drivers and accidents: A meta analysis and data mining application on traffic accident data. Expert Systems with Applications. 2005;

BEZIE, Y. et al. Therapeutic compliance: a prospective analysis of various factors involved in the adherence rate in type 2 diabetes. Diabetes and Metabolism, v. 32, s./n., p. 611-6, 2006.

BOENTE, Alfredo Nazareno Pereira ; GOLDSCHMIDT ; ESTRELA, Vania Vieira . Uma Metodologia de Suporte ao Processo de Descoberta de Conhecimento em Bases de Dados. In: V Simpósio de Excelência em Gestão e Tecnologia, 2008, Resende - RJ. V SEGeT, 2008. v. 1. p. 4-5.

BOSI, P. L. et al. Prevalência de diabete melito e tolerância à glicose diminuída na população urbana de 30 a 79 anos da cidade de São Carlos, São Paulo. Arquivo Brasileiro de Endocrinologia e Metabologia, v. 53, n. 6, p. 726-32, 2009.

BRASIL. Ministério da Saúde. Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil 2011-2022. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação de Saúde. Brasília: Ministério da Saúde, 2011. 148 p.

BRASIL. Ministério da Saúde. Cadernos de Atenção Básica - n. 16, Serie A. Normas e Manuais Técnicos. Brasília: Secretaria de Atenção à Saúde, Ministério da Saúde, 2006. 64 p.

BRASIL. Ministério da Saúde. Organização Pan-Americana de Saúde. Avaliação do Plano de Reorganização da Atenção à Hipertensão arterial e ao Diabetes mellitus no Brasil. Brasília: Ministério da Saúde, 2004. 64p.

BRASIL. Ministério da Saúde. Secretaria de Gestão Estratégica e Participativa. Departamento de Apoio à Gestão Participativa. Caderno de educação Popular e Saúde. Brasília: Ministério da Saúde, 2007, 80 p.

BRASIL. Ministério da Saúde. Secretaria-Executiva. Área de Economia da Saúde e Desenvolvimento. Avaliação de tecnologias em saúde: ferramentas para a gestão do SUS / Ministério da Saúde, Secretaria-Executiva, Área de Economia da Saúde e Desenvolvimento. – Brasília : Editora do Ministério da Saúde, 2009. 110 p. : il. – (Série A. Normas e Manuais Técnicos).

BRUSSO, M. J. Access Miner: Uma proposta para a Extração de Regras de Associação Aplicada à Mineração do Uso da Web. Master's thesis, PPGC da UFRGS, Porto Alegre - RS, 2000

CABRAL-DE-OLIVEIRA, Antonio Cesar; RAMOS, Plínio Santos e ARAUJO, Claudio Gil Soares de. Distância do domicílio ao local de exercício físico não influenciou a aderência de 796 participantes. Arq. Bras. Cardiol. [online]. 2012

CAMILO, C. O.; SILVA, J. C. Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas. Instituto de Informática, Universidade Federal de Goiás, Goiânia, 2009, p. 29.

CAMPOS, J. J.; ROS, J. A. C.; ANGUERA, J. M. H.; CANAL, D. J.; TORELLÓ, J. M.P. La supresión de La educación especializada empeora El control metabólico em diabetes tipo 2. Atención Primaria, v. 41, n. 12, p. 681-687, 2009.

CARDOSO, Olinda N. P., MACHADO, Rosa T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. Revista de Administração Pública. Rio de Janeiro 42(3) : 495-528, Maio/Jun. 2008.

CAROLINO, I. D. R. et al. Fatores de risco em pacientes com Diabetes mellitus tipo 2. Revista Latino-americana de Enfermagem, v. 16, n. 2, p. 238-44, 2008.

CHIARA, Ramon. Aplicação de Técnicas de Data Mining em Logs de Servidores Web. 2003. Dissertação (Mestrado). Instituto de Ciências Matemáticas e de Computação - ICMC/USP. 2003.

COMIOTTO, G.; MARTINS, J. J. Promovendo o autocuidado ao indivíduo portador de diabetes: da hospitalização ao domicílio. Arquivos Catarinenses de Medicina, v. 35, n. 3, p. 59-64, 2006.

COSSON ICO, NEY-OLIVEIRA F, ADAN LF. Avaliação do conhecimento de medidas preventivas do pé diabético em pacientes de Rio Branco, Acre. Arq Bras Endocrinol Metab. 2005;49(4):548-56.

COSTA JA, BALGA RSM, ALFENAS RTG, COTTA RMM. Promoção da saúde e diabetes: discutindo a adesão e a motivação de indivíduos diabéticos participantes de programas de saúde. CienSaude Colet 2011.

DALLAGASSA, Marcelo Rosano. Concepção de uma metodologia para identificação de beneficiários com indicativos de diabetes melitus tipo 2. 2009.105 f. Dissertação (Mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba.

DAVIES, M. J. et. al. Effectiveness of the diabetes education and self management for ongoing and newly diagnosed (DESMOND) programme for people with newly diagnosed type 2 diabetes : cluster randomised controlled trial. BMJ, v 336, p. 491- 495, 2008.

DEVORE, J. L. Probabilidade e estatística: para engenharia e ciências. São Paulo: Pioneira Thomson Learning, 2006.

DIAS, M. M. Um modelo de formalização do processo de sistema de descoberta de Conhecimento em banco de dados. 2001. Tese (Doutorado)-Pós Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis, Santa Catarina, 2001.

DIAS, Maxwell M.; et all. Aplicação de Técnicas de Mineração de Dados no Processo de Aprendizagem na Educação a Distância. XIX Simpósio Brasileiro de Informática na Educação. Florianópolis, 2008.

DIB, Sérgio Atala; CHACRA, Antonio Roberto. Diabete Melito Tipo 1. In: SAAD, Mário J. A.; MACIEL, Rui M. B.; MENDONÇA, Berenice B. Endocrinologia. São Paulo: Atheneu, 2007. Cap. 39, p. 705-737.

DIEDRICH, A.; MUNROE, D. J.; ROMANO, M. Promoting physical activity for persons with diabetes. The Diabetes Educator, v. 36, n. 1, p. 132-140, 2010.

DIRETRIZES DA SOCIEDADE BRASILEIRA DE DIABETES: 2013-2014/Sociedade Brasileira de Diabetes ; [organização José Egidio Paulo de Oliveira, Sérgio Vencio]. – São Paulo: AC Farmacêutica, 2014.

DISHMAN, R. K. Advances in exercise adherence. Champaign: Human Kinetics, 1994.

Dissertação (Mestrado em Enfermagem) – Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

DULLIUS J. Educação em Diabetes por meio de programa orientado de atividades físicas [tese de doutorado]. Brasília (DF). Faculdades de Ciências da Saúde – UnB; 2003.

DULLIUS, J. & cols. Diabetes Mellitus: Saúde, Educação, Atividades Físicas. Brasília: Editora UnB. 2007; 1º Ed. p 72-79.

ELDER J. e PREGIBON, D. A Statistical Perspective on Knowledge Discovery in Databases. Pages 83-116. In (Fayyad et al., 1996), 1996.

FARIA, Heloisa T.G. et al. Qualidade de vida de pacientes com diabetes mellitus antes e após participação em programa educativo. *Rev. esc. enferm. USP*, São Paulo, v. 47, n. 2, p. 348-353, 2013

FAYYAD, USAMA; PIATETSKI-SHAPIRO, GREGORY; SMYTH, PADHRAIC. The KDD process for extracting useful knowledge from volumes of data. *Communications Of The Acm*, New York, v. 39, n. 11, p. 27-34, 1996.

FECHIO, J. J.; MALERBI, F. E. K. Adesão a um programa de atividade física em adultos portadores de diabetes. *Arquivos Brasileiros de Endocrinologia e Metabologia*, v. 48, n. 2, p. 267-275, 2004.

FERREIRA CLRA, FERREIRA MG. Características epidemiológicas de pacientes diabéticos da rede pública de saúde – análise a partir do sistema HiperDia. *Arq Bras Endocrinol Metab*. 2009; 53(1): 80-6.

FERREIRA, JOSÉ ALVES. Data Mining em Banco de Dados de Eletrocardiograma. Tese de Doutorado- Instituto Dante Pazzanese de Cardiologia. Universidade de São Paulo. São Paulo, 2014.

FIGUEIREDO MFS, Rodrigues-Neto JF, Leite MTS. Modelos aplicados às atividades de educação em saúde. *Rev Bras Enferm*. 2010;63(1):117-21

FRANCIONI, F. F.; SILVA, D. G. V. O processo de viver saudável de pessoas com diabetes mellitus através de um grupo de convivência. *Texto e Contexto Enfermagem*, v. 16, n. 1, p. 105-11, 2007.

FREITAS, R. W. J. F. et al. Fatores associados à não adesão aos agentes antidiabéticos orais em pacientes com diabetes mellitus tipo 2. *Revista de la ALAD*, v. 1, n. 4, p. 170-85, 2011.

FUNNELL, M.M., et.al. . (2010). National Standards for Diabetes Self- Management Education. *Diabetes Care*, 33(suppl.1), 89-96. 2010.

GALVÃO, Noemi Dreyer. Aplicação da Mineração de Dados em bancos da segurança e saúde pública em acidentes de transporte. Tese (Doutorado) Universidade Federal de São Paulo. Escola Paulista de Medicina - São Paulo. 2009.

GALVÃO, Noemi Dreyer; MARIN, Heimar de Fátima. Técnica de Mineração de Dados: uma revisão da literatura. *Actapaul. enferm.*, São Paulo, v. 22, n. 5, Oct. 2009.

GILL, J.M., & COOPER, A.R. (2008). Physical Activity and Prevention of Type 2 Diabetes Mellitus. *Sports Med*, 38(10), 807-824.

GIMENES, H. T.; ZANETTI, M. L.; HAAS, V. J. Fatores relacionados à adesão do paciente diabético à terapêutica medicamentosa. *Revista Latino-americana de enfermagem [online]*, v. 17, n. 1, p. 46-51, 2009.

GLEESON-KREIG, J. Social Support and Physical Activity in type 2 diabetes: a social ecologic approach. *The Diabetes Educator*, v. 34, n. 6, p. 1037-1044, 2008.

GOLDSCHMIDT R, Passos E. *Data Mining: Guia Prático. Conceitos, técnicas, ferramentas, orientações e aplicações*. Rio de Janeiro: Elsevier; 2005.

GOMES HM, Haugt LG, Carvalho DR. *Mineração de Dados Temporal: Descobertas de Regras De Causa e Efeito*. In: *Anais do V Congresso Sul Brasileiro de Computação*; 2010

GOMES, M. B. et al. Prevalência de sobrepeso e obesidade em pacientes com diabetes mellitus do tipo 2 no Brasil: estudo multicêntrico nacional. *Arquivo Brasileiro de Endocrinologia e Metabologia*, v. 50, n. 1, p. 136-44, 2006.

GOWER, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 77, 623-637

GREENSPAN, Francis S.; GARDNER, David G. *Endocrinologia Básica e Clínica*. 7. ed. Rio de Janeiro: Mcgraw-hill, 2006. Cap. 17, p. 541-613.

GRILLO, M. F. F.; GORINI, M. I. P. C. Caracterização de pessoas com Diabetes mellitus tipo 2. *Revista Brasileira de Enfermagem*, v. 60, n. 1, p. 49-54, 2007.

GROSSI SAA. O manejo do diabetes mellitus sob a perspectiva da mudança comportamental. In: *Sociedade Brasileira de Diabetes (SBD). Cuidados de enfermagem em diabetes mellitus*. São Paulo; 2009.

GUZMÁN, J. R. ET AL. Treatment of type 2 diabetes in Latin America: a consensus statement by the medical associations of 17 Latin American countries. *RevistaPanamericana de SaúdePública*, v. 28, n. 6, p. 463-71, 2010.

HAMERLY, Greg; ELKAN, Charles. Learning the k in k-means. In: *conference on neural information processing systems (NIPS)*, 17., 2012, San Diego. *Anais...* San Diego: University of California, Department of Computer Science and Engineering, 2012.

HAN, J., KAMBER, M., PEI, J. *Data Mining: Concepts and Techniques*. 3. ed. San Francisco: Morgan Kaufmann, 2011.

INTERNACIONAL DIABETES FEDERATION- IDF Diabetes atlas 2014. Disponível em: [HTTP://WWW.DIABETES.ORG.BR/IMAGES/PDF/ATLAS-IDF-2014.PDF](http://www.diabetes.org.br/images/pdf/atlas-idf-2014.pdf). Acesso em 15 de maio de 2015.

JOHNSON, R. A.; WICHERN, D. W. *Aplied multivariate statistical analysis*. 4th, ed. Prentice-Hall, New Jersey, 1999.

KAMPPFF , Adriana J. C.. *Mineração de Dados Educacionais para Geração de Alertas em Ambientes Virtuais de Aprendizagem como Apoio à Prática Docente*. 2009. 189 f. Tese (Doutorado em Informática na Educação) – UFRGS, Universidade Federal do Rio Grande do Sul, 2009.

KAUR H, WASAN SK. Empirical study on applications of Data Mining techniques in healthcare. *Journal of Computer Science*. 2006; 2(2):194-200.

KOBUS, Luciana Schleder Gonçalves. Aplicação da descoberta de conhecimento para identificação de usuários com doenças cardiovasculares elegíveis para programas de gerenciamento de caso. 2006. 145f. Dissertação (Mestrado) - Pontifícia Universidade Católica do Paraná, Curitiba.

LAHIRI, R. Comparison of data mining and statistical techniques for classification model. [S.l.]: Thesis of Master of Science at the Louisiana State University, US, 2006

LEE, I. N.; LIAO, S. C.; EMBRECHTS, M. Data mining techniques applied to medical
LEITE, Cícilia Raquel Maia. DATA MINING “Mineração de Dados e suas Descobertas”. Mossoró-RN: Universidade do Estado do Rio Grande do Norte. 2002.

LEITE, M.M.J.; PRADO, C.; PERES, H.H.C. Educação em saúde: desafios para uma prática inovadora. 1. ed. São Caetano do Sul, SP: Difusão Editora, 2010.

LEITE, S. A. O. et al. Enfoque multidisciplinar ao paciente diabético: avaliação do impacto do “Staged diabetes management” em um sistema de saúde privado. *Arquivo Brasileiro de Endocrinologia e Metabologia*, v. 45, n. 5, p. 481-6, 2001.

LEITE, S. A. O.; ZANIM, L. M.; GRANZOTTO, P. C. D.; LAMOUNIER, R. N. Pontos básicos de um programa de educação ao paciente com diabetes mellitus tipo 1. *Arquivos Brasileiros de Endocrinologia e Metabologia*, v. 52, n. 2, p. 233-242, 2008

LI, R. ET AL. Cost-effectiveness of interventions to prevent and control diabetes mellitus: a systematic review. *Diabetes care*, v. 33, n. 8, p. 1872-94, 2010.

LOPES EM, Anjos SJSB, Pinheiro AKB. Tendência das ações de educação em saúde realizadas por enfermeiros no Brasil. *RevEnferm UERJ*. 2009;17 (2):273-7.

LUCAS, Anelise de Macedo. Utilização de Técnicas de Mineração de Dados considerando os Aspectos Temporais. 2002. Dissertação (Mestrado). Porto Alegre: PPGC da UFRGS, 2002.

LYRA, R. et al. Prevalência de diabetes mellitus e fatores associados em população urbana adulta de baixa escolaridade e renda do sertão nordestino brasileiro. *Arquivo Brasileiro de Endocrinologia e Metabologia*, v. 54, n. 6, p. 560-6, 2010.

MACEDO, D. C. e Matos, S. N., Extração de Conhecimento Através da Mineração de Dados, *Revista de Engenharia e Tecnologia*. v. 2, n. 2, p.22-30, 2010.

MACHADO FILHO, OTTO MOURA Ambiente de Mineração de Dados Utilizando Redes Neurais Otimizadas por Algoritmos Genéticos e Técnica de Visualização [Rio de Janeiro] Tese - Universidade Federal do Rio de Janeiro, COPPE, 2006.

MADIGAN EA, CURET OL, ZRINYI M. Workforce analysis using data mining and linear regression to understand HIV/AIDS prevalence patterns. *Human Resources for Health*. 2008;6:2.

MANTOVANI MF, Maciel KF, Pelinski A, Gaio DM, Fusuma F, Ulbrich E. Dificuldades no tratamento da doença crônica: relato de Experiência de atividade de extensão. Ciênc. Cuid. Saúde [on-line]. 2011. Disponível em: URL: <http://periodicos.uem.br/ojs/index.php/ciencucuidsaude/article/view/10842>

MATHIEU ME, BROCHU M, BELIVEAU L. DIABETACTION: changes in physical activity practice, fitness, and metabolic syndrome in type 2 diabetic and at-risk individuals. Clin J Sport Med. 2008; 18 (1): 70-5.

MATTOS MC, PELEGRIN DC, CASAGRANDE DP, SIMÕES PWTA, CHARNOVSKI R, BETTIOL J. Análise de Alergia e Rinite em Escolares por meio da Shell Orion Data Mining Engine. Anais do Congresso da SBIS; 2006; Florianópolis / SC.

MAULDON, M.; MELKUS, G. D.; CAGGANELLO, M. Tomando control: a culturally appropriate diabetes education program for Spanish-speaking individuals with type 2 diabetes mellitus – evaluation of a pilot project. The Diabetes Educator, v. 32, n. 5, p. 751-760, 2006.

MCNICHOLAS, P.; MURPHY, T.; OREGAN, M. Standardising the lift of an association rule. Computational Statistics & Data Analysis, [S.l.], v.52, n.10, p.4712–4721, June 2008.

MENDES, R; SOUSA, N; BARATA, J.L. Physical activity and public health: recommendations for exercise prescription. Acta Med Port; 24(6): 1025-30, 2011.

MILLER, S. T.; MAROLEN, K. N.; BEECH, B. M. Perceptions of Physical Activity and Motivational Interviewing Among Rural African American Women with Type II Diabetes. Womens Health Issues, v. 20, n. 1, p. 43-49, 2010.

MINNIE, D; Srinivasan S., "Application of Knowledge Discovery in Database to Blood Cell Counter Data to Improve Quality Control in Clinical Pathology", 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications, pp.338-342, 2011.

MODENEZE DM. Qualidade de vida de portadores de Diabetes Mellitus tipo II frente às oscilações no nível de atividade física, aspectos socioeconômicos e presença de comorbidades [tese]. Campinas: Universidade Estadual de Campinas; 2011.

MORAES, A.B.A., Rolim, G.S. & Costa Jr., A.L. (2009). O processo de adesão numa perspectiva analítico comportamental. Rev. bras. ter. comport. cogn., 11(2): 329-345. http://pepsic.homolog.bvsalud.org/scielo.php?script=sci_arttext&pid=S1517-55452009000200009&lng=pt&nrm=isso

MORAES, S. A. et al. Prevalência de diabetes mellitus e identificação de fatores associados em adultos residentes em área urbana de Ribeirão Preto, São Paulo, Brasil, 2006: Projeto OBEDIARP. Caderno de Saúde Pública, v. 26, n. 5, p. 929-41, 2010.

MOURA, K.H.S. Mineração de Dados Petrolífero Utilizando a Ferramenta WEKA. Monografia – UERN. Mossoró, 2007.

MUSACCHIO, N. et al. Impact of a chronic care model based on patient empowerment on the management of type 2 diabetes: effects of the SINERGIA programme. *Diabetic Medicine*, v. 28, s./n., p. 724-30, 2011.

NASCIMENTO, L. S.; GUTIERREZ, M. G. R.; DE DOMENICO, E. B. L. Programas educativos baseados no autogerenciamento: uma revisão integrativa. *Revista Gaúcha de Enfermagem*, v. 31, n. 2, p. 375-82, 2010.

NEGRI, C.; BACCHI, E.; MORGANTE, S.; SOAVE, D.; MARQUES, A.; MENGHINI, E.; MUGGEO, M.; BONORA, E.; MOGHETTI, P. Supervised Walking Groups to Increase Physical Activity in Type 2 Diabetic Patients. *Diabetes Care*, v. 33, n. 11, p. 2333-2335, 2010.

OMS. Envelhecimento ativo: um Projeto de Saúde Pública. In: *Anais do 2º Encontro Mundial das Nações Unidas sobre Envelhecimento*. Madri, p. XX, 2002.

OMS. Organização Mundial de Saúde (2003). *Cuidados inovadores para condições crônicas: componentes estruturais de ação: relatório mundial*. Tradução de documento internacional. Brasília, DF.

OTERO, L. M.; ZANETTI, M. L.; TEIXEIRA, C. R. S. Características sociodemográficas e clínicas de portadores de diabetes em um serviço de atenção básica à saúde. *Revista Latino-americana de Enfermagem*, v. 15, n. especial, p. 768-73, 2008.

PAIVA, L. C. F.; COSTA, J. S. D.; JARDIM, V. M. R.; SOARES, M. C.; D'AVILA, A. D. Avaliação da efetividade dos cuidados de saúde nos pacientes com Diabetes Mellitus em uma comunidade de Pelotas, RS: processo e resultado. *Revista da Associação Médica do Rio Grande do Sul - AMRIGS*, v. 48, n. 1, p. 5-10, jan.mar. 2004.

PASTA, ARQUELAU. Aplicação da técnica de data mining na base de dados do ambiente de gestão educacional: um estudo de caso de uma instituição de ensino superior de blumenau-sc. Dissertação de mestrado. 2011

PFAFF M. et al . Prediction of Cardiovascular Risk in Hemodialysis Patients by Data Mining. *Methods of Information in Medicine*. 2004.

PHILIS-TSIMIKAS, A. et al. Peer-led diabetes education programs in high-risk mexican americans improve glycemic control compared with standard approaches. *Diabetes Care*, v. 34, s./n., p. 1926-31, 2011.

PIMAZONI NETTO, A. (2008). *Manual básico sobre diabetes*. Grupo de Educação e Controle do Diabetes, Centro Integrado de Hipertensão e Metabologia Cardiovascular, Hospital do Rim e Hipertensão da Universidade Federal de São Paulo, UNIFESP

PONTIERI, F. M.; BACHION, M. M. Crenças de pacientes diabéticos acerca da terapia nutricional e sua influência na adesão ao tratamento. *Ciência & Saúde Coletiva*, v. 15, n. 1, p. 151-60, 2010.

QUEIROZ FA, PACE AE, SANTOS CB. Adaptação cultural e validação do instrumento Diabetes-39 (D-39): versão para brasileiros com diabetes mellitus tipo 2 - fase 1. *Rev Latino.Am.Enferm.*2009.

QUONIAM L, TARAPANOFF K, ARAUJO JUNIOR RH, ALVARES L. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. *Ci Inf.* 2001; 30(2):20-28.

R Development Core Team. R: A language and environment for statistical computing. R. Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>, 2006.

RÊGO MAB, NAKATANI AYK, BACHION MM. Educação para a saúde como estratégia de intervenção de enfermagem às pessoas portadoras de diabetes. *Rev Gaúcha Enferm.* 2006.

REZENDE, S. O. **Sistemas Inteligentes**: fundamentos e aplicações. Barueri, SP: RIBAS, C.R.P., Santos, M.A., Teixeira, C.R.S. & Zanetti, M.L. (2009). Expectativas de mulheres com diabetes em relação a um programa de educação em saúde. *Revista Enfermagem UERJ*, 17(2), 203-208.

RICKHEIM PR, Weaver TW, Flader JL, Kendall DM. Assessment of group versus individual diabetes education. *Diabetes Care.* 2002;25:269-74.

ROBISON, J.I; ROGERS, M.A. Adherence to exercise programmes. *Sports Medicine, Pensilvânia*, v.17, p. 39-52, 1994.

ROCHA, C. A. J. Redes Bayesianas Para Extração de Conhecimento de Bases de Dados, Considerando a Incorporação de Conhecimento de Fundo e o Tratamento de Dados Incompletos. Dissertação de Mestrado, ICMC - Universidade de São Paulo, 1999.

ROUSSEEUW, P.J. (1986). Silhouettes: A graphical and to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20, 53-65.

RUBIN O, Azzolin K, Muller S. Adesão ao tratamento de diabetes mellitus tipo 1 atendidos em um programa especializado em Porto Alegre. *Medicina.* 2011;44(4):367-76.

SAAD, Mário J. A. Classificação do Diabete Melito e Critérios Diagnósticos. *Endocrinologia.* São Paulo: Atheneu, 2007. Cap. 38, p. 699-703.

SAMPAIO FAA, Melo RP, Rolim ILTP, Siqueira RC, Ximenes LB, Lopes MVO. Avaliação do comportamento de promoção da saúde em portadores de Diabetes Mellitus. *Acta Paul Enferm.* 2008; 21(1): 84-8.

SANTOS, L. M. Competências dos Profissionais de Saúde nas Práticas Educativas em Diabetes Tipo 2 na Atenção Primária em Saúde. 2011. 81 f. Dissertação (Mestrado em Enfermagem) – Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

- SCHILLINGER, D. et al. Examining the reach of diabetes self-management support, 2007.
- SCHMIDT, M. A. et al. Prevalência de diabetes e hipertensão no Brasil baseada em inquérito de morbidade autorreferida, Brasil, 2006. *Revista de Saúde Pública*, v. 43, Suplemento (2), p. 74-82, 2009.
- SCOSS, Anne M.. A Clusterização e Classificação no Processo De Data Mining para Análise do Desempenho Docente no Ensino de Graduação. 2006. 86 f. Trabalho de Conclusão de Curso (Especialização) - Universidade do Extremo Sul Catarinense - UNESC, Criciúma, 2006.
- SELLI, L.; PAPALÉO, L. K.; MENEGHEL, S. N.; TORNEROS, J. Z. Técnicas educacionales en el tratamiento de la diabetes. *Cadernos de Saúde Pública*, v. 21, n. 5, p. 1366-1372, 2005.
- SHAMA A. Mansotra V. Emerging applications of data mining for healthcare management - A critical review. 2014 International Conference on Computing for Sustainable Global Development (INDIACom), New Dehli. pages 377-382. 2014.
- SHARMA, S.; OSEI-BRYSON, K.-M.; KASPER, G. M. Evaluation of an integrated Knowledge Discovery and Data Mining process model. *Expert Systems with Applications*, [S.l.], v.39, p.11335–11348, Feb. 2012.
- SHAW L, Chan F, McMahon B. Intersectionality and disability harassment: the interactive effects of disability, race, age and gender. *RCB*. 2012;55(2):82-91.
- SIGAL RJ, FISHER SJ, MANZON A, MORAIS JA, HALTER JB, VRANIC M, MARLISS EB: Glucoregulation during and after intense exercise: effects of alpha-adrenergic blockade. *Metabolism* 49:386–394, 2000
- SILVA, Livia Oliveira. Conhecimento dos portadores de Diabetes Mellitus do tipo 1 em relação ao autocuidado, no município de Paracatu/MG. 2010/ Livia Oliveira Silva; orientador: Iris Leda Nery Ferreira – 2010
- SILVA, Luís Alexandre Estevão da Metodologia Baseada em Mineração de Dados Aplicada na Análise da Distribuição Espacial de Agrupamentos de Espécies Vegetais / Luís Alexandre Estevão da Silva. – Rio de Janeiro: UFRJ/COPPE, 2013.
- SOCIEDADE BRASILEIRA DE CARDIOLOGIA / SOCIEDADE BRASILEIRA DE HIPERTENSÃO / SOCIEDADE BRASILEIRA DE NEFROLOGIA. VI Diretrizes Brasileiras de Hipertensão. *ArqBrasCardiol* 2010; 95(1 supl.1): 1-51
- SOCIEDADE BRASILEIRA DE DIABETES (2007). Tratamento e acompanhamento do diabetes melitus. Diretrizes da Sociedade Brasileira de Diabetes. http://www.diabetes.org.br/educacao/docs/Diretrizes_SBD_2007.pdf

SOCIEDADE BRASILEIRA DE DIABETES. Algoritmo para tratamento do diabetes tipo 2. Atualização 2009. Posicionamento oficial no 2. 2009. 26 p.

SOCIEDADE BRASILEIRA DE DIABETES. Como prescrever exercício no tratamento do diabetes mellitus: Diretrizes da Sociedade Brasileira de Diabetes 2009. 3. ed. São Paulo: SBD, 2009. Disponível em: <<http://www.diabetes.org.br/>>. Acesso em: 15 jun. 2015.

SOCIEDADE BRASILEIRA DE DIABETES. Consenso brasileiro sobre diabetes. Diretrizes da Sociedade Brasileira de Diabetes: 2013-2014. São Paulo: AC Farmacêutica, 2014.

SOCIEDADE BRASILEIRA DE DIABETES. Diretrizes SBD 2007. Tratamento e acompanhamento do diabetes Mellitus Disponível em: <HTTP://WWW.DIABETES.ORG.BR/EDUCAÇÃO/docs/Diretrizes_SBD_2007.pdf> Acesso em: 14 de maio. 2015.

STEIN JUNIOR, Altair Von. Descoberta de regras por meio de KDD para a classificação de micro áreas homogêneas de risco. 2008. 106f. Dissertação (Mestrado)- Pontifícia Universidade Católica do Paraná, Curitiba.

STEINER MTA, SOMA NY, SHIMIZU T, NIEVOLA JC, NETO-STEINER PJ. Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. Gest & Prod. 2006; 13(2):325-337.

STEWART, G. L. et al. Control of type 2 diabetes mellitus among general practitioners in private practice in nine countries of Latin America. Revista Panamericana de Salud Pública, v. 22, n. 1, p. 12-20, 2007.

TAN, P. N., STEINBACH, M., KUMAR, V. Introdução ao Data Mining. 1st. ed. Rio de Janeiro: Editora Ciência Moderna, 2009.

TEIXEIRA CRS, Zanetti ML, Martins DC. Parâmetros clínicos de pacientes diabéticos atendidos no programa educativo do serviço de medicina preventiva- sempre. CiêncCuidSaúde 2005; 4(3): 231-41.

TORQUATO, M. T. C. G.; MONTENEGRO, J. R. M.; VIANA, L. A.; SOUZA, R. A.; LANNA, C. M.; LUCAS, J. C.; BIDURIN, C.; FOSS, M. C. Prevalence of diabetes mellitus and impaired glucose tolerance in the urban population aged 30-69 years in Ribeirão Preto (São Paulo), Brazil. São Paulo Medical Journal, São Paulo, v. 121, n. 6, p. 224-230, 2003.

TORRES HC, Roque C, Nunes C. Visita domiciliar: Estratégia educativa para o autocuidado de clientes diabéticos na atenção básica. Rev. enferm. UERJ. 2011; 19(1): 89-93.

TORRES, H. C. Avaliação de um programa educativo para controle de diabetes mellitus tipo 2 em Belo Horizonte, MG. 115p. Tese - Escola Nacional de Saúde Pública, FIOCRUZ, São Paulo, 2004.

TORRES, H. C.; AMARAL, M. A.; AMORIM, M. M.; CYRINO, A. P.; BODSTEIN, R. Capacitação de profissionais da atenção primária à saúde para educação em Diabetes Mellitus. *Acta Paulista de Enfermagem*, v. 23, n. 6, p. 751-756, 2010.

TORRES, H. C.; PACE, A. E.; STRADIOTO, M. A. Análise sociodemográfica e clínica de indivíduos com diabetes tipo 2 e sua relação com o autocuidado. *Cogitare Enfermagem*, v. 15, n. 1, p. 48-54, 2010.

TORRES, Heloisa de Carvalho et al. Avaliação estratégica de educação em grupo e individual no programa educativo em diabetes. *Rev. Saúde Pública*, São Paulo, v. 43, n.2, p.291-298, Apr.2009. Acesso em 30 mar. 2015. [HTTP://DX.DOI.ORG/10.1590/S0034-89102009005000001](http://dx.doi.org/10.1590/S0034-89102009005000001).

TRINDADE, Carla Machado et al. Technology in health: knowledge discovery in public health databases: study of viral hepatitis in the state of Paraná, Brazil. *Iberoamerican Journal of Applied Computing*, Ponta Grossa, v. 2, n. 2, 2012.

TRIOLA, M.F. Introdução à estatística. 9. ed. Rio de Janeiro: LTC, 2005.

VALE, M. N. Agrupamentos de dados: Avaliação de Métodos e Desenvolvimento de Aplicativo para Análise de Grupos. Rio de Janeiro, 2005. 120 f. Dissertação (Mestrado) – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro

VANCEA, D. M. M. et al. Effect of frequency of physical exercise on glycemic control and body composition in type 2 diabetic patients. *Arquivos brasileiros de cardiologia*, Rio de Janeiro, v. 92, n. 1, p. 22-28, 2009.

VANCINI; R.L.; Lira, C.A.B. Aspectos Gerais do Diabetes Mellitus e Exercício Centro de Estudos de fisiologia do Exercício UNIFESP. 2004.

VIANNA, Rossana. Mineração de dados e características da mortalidade infantil. *CadernoSaúdePública*, Rio de Janeiro, v. 26, n. 3, p. 535-542, 2010.

VIEIRA, Gisele de Lacerda Chaves. Avaliação da educação em grupos operativos com usuários diabéticos tipo 2 em Unidades Básicas de Saúde – Belo Horizonte – MG. 2011. 124f.

WHO. Package of essential non communicable disease interventions for primary health care in low-resource settings. Geneva: World Health Organization, 2010.

WITTEN, I. H.; FRANK, E.; HALL, M. A. DATA MINING. Practical Machine Learning Tools and Techniques.-3rd ed. [S.l.]: Morgan Kaufmann, 2011.

WU, X., KUMAR, V., QUINLAN, J. R., GHOSH, J., YANG, Q., MOTODA, H., MCLACHLAN, G. F., NG, A., LIU, B., YU, P. S., ZHOU, Z., STEINBACH, M., HAND, D. J., S. D. “Top 10 algorithms in data mining”, *Knowledge and Information Systems*, v. 14, n. 1, pp. 1–37, Dez. 2007.

YAMADA, Aparecida Teruko Tominaga, et.al. Manual de orientação clínica: diabetes mellitus / Clinical orientation manual: diabetes mellitus. São Paulo; SES/SP; 2011. 46 p. ilustrado.

ZANETTI, M. L. et. al. Satisfacción del diabético em seguimiento em um programa de educación em diabetes. Revista Latino-americana de enfermagem, v. 15, n. 4, 2007.

ANEXO A – PUBLICAÇÕES

1. Participação em congressos:

- 1.1. 10th International Conference on Mathematical Problems in Engineering, Aerospace and Sciences, to be held at the Narvik University, Norway, during July 15-18, 2014. Entitled: Analysis of the pressure data of autistic walk.
- 1.2. 5th European Conference on Mechanism Science (EUCOMES 2014), realizado em Guimarães, Portugal, 16-20 de setembro de 2014. “Characterization and Analysis of Data in Autistic Walk Pressure”.
- 1.3. Evidências sobre a adesão ao tratamento do Diabetes Mellitus e sua relação com a qualidade de vida: uma revisão integrativa. Autores: Thiane Marques Torquato, Suélia de Siqueira Rodrigues Fleury Rosa, Mário Fabricio Fleury Rosa, Thiane Marques Torquato. Submetido ao XX Congresso da Sociedade Brasileira de Diabetes- Porto Alegre – RS.

1. Submissão de manuscritos:

- 1.1. Expectativas da utilização do biomaterial látex para o tratamento do pé diabético: revisão integrativa. Autores: Mário Fabricio Fleury Rosa; Thiane Marques Torquato; Suélia Rodrigues Fleury Rosa;
- 1.2. Seringas descartáveis para Insulinoterapia: porque reutilizar se são descartáveis? Sessão: Comentários. Autores: Thiane Marques Torquato, Thiane Marques Torquato, Walterlânia Silva Santos.
- 1.3. Programa de Educação em Diabetes com Ênfase em Exercícios Físicos: Análise e Mineração de Dados. Thiane Marques Torquato; Suélia Rodrigues Fleury Rosa; Juliana de Faria Fracon e Romão. Submetido a Revista Brasileira de Atividade Física & Saúde.

2. Artigo aceito para publicação:

2.1. Characterization and Analysis of Data in Autistic Walk Pressure. Springer, 2014.

Autores: SOUZA, L. L. ; CARVALHO, M. O. M. ; ROSA, S. S. R. F. ; TORQUATO, T. M.

2.2. Úlceras Por Pressão: Proposta de um Colchão Inteligente Derivado Do Látex Natural

(Hevea Brasiliensis). Autores: Thiane Marques Torquato, Suélia Rodrigues Fleury Rosa, Mário Fabricio Fleury Rosa, Maria Joana de Carvalho, Roozbeh Tahmasebi.

Submetido a revista: Ciência e Engenharia.

3. Capítulo de livro publicado.

3.1. Characterization and Analysis of Data in Autistic Walk Pressure. In: 5th European

Conference on Mechanism Science, 2014, Guimarães. New Trends in Mechanism and Machine Science - From Fundamentals to Industrial Applications. Zug: Springer

International Publishing AG, 2014. v. 24. Autores: SOUZA, L. L. ; CARVALHO, M. O. M. ; ROSA, S. S. R. F. ; TORQUATO, T. M.