

O uso da metodologia CRISP-DM para apoiar a análise de dados no aplicativo *mHealth* IUProst

Noelí Antonia Pimentel Vaz^{1,2}, Deborah Silva Alves Fernandes¹, Sergio T. Carvalho¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74001-970 – Goiânia – GO – Brasil

²Instituto de Ciências Exatas e Tecnológicas – Universidade Estadual de Goiás (UEG)
Caixa Postal 459 – 75220-400 – Anápolis – GO – Brasil

noelivaz@discente.ufg.br, deborah@inf.ufg.br, sergiocarvalho@ufg.br

Abstract. *The use of mobile devices in healthcare aims to assist in the control and monitoring of diseases and clinical conditions. Mobile health applications (mHealth) contribute to individuals' self-care. IUProst is a mHealth application that aids patients during the treatment of urinary incontinence and comorbidity resulting from prostate removal surgery in cancer patients. Despite the potential of mHealth applications like IUProst to assist in the treatment of urinary incontinence, the low adherence of users underscores the urgency of implementing effective engagement mechanisms. The objective of this article is to report on a study that employs the CRISP-DM methodology to identify patterns, trends, and insights in IUProst data, aiming to uncover demands to support the application's future development cycles. Analyses of the results obtained in the business understanding and data understanding phases reveal a significant number of users and exercises performed but with low participation in the proposed cognitive behavioral treatment.*

Resumo. *O uso de dispositivos móveis na saúde visa auxiliar no controle e monitoramento de doenças e condições clínicas, colaborando com o autocuidado das pessoas. Nesse contexto, o IUProst destaca-se como uma aplicação mHealth que auxilia pacientes durante o tratamento da incontinência urinária, uma comorbidade decorrente da cirurgia de retirada de próstata em pacientes que enfrentaram o câncer. Apesar do potencial das aplicações mHealth, como o IUProst, para auxiliar no tratamento da incontinência urinária, a baixa adesão dos usuários destaca a urgência de implementar mecanismos de engajamento eficazes. O objetivo deste artigo é relatar uma pesquisa que utiliza a metodologia CRISP-DM para identificar padrões, tendências e percepções nos dados do IUProst, visando levantar demandas para subsidiar os próximos ciclos de evolução e o aprimoramento da aplicação. Análises dos resultados obtidos nas fases de compreensão do negócio e compreensão dos dados, revelaram um número significativo de usuários e exercícios realizados, porém com baixa participação no tratamento cognitivo comportamental proposto.*

1. Introdução

O uso de dispositivos móveis na área da saúde visa auxiliar no controle e monitoramento de doenças e condições clínicas, por meio das aplicações de saúde móvel. Conhecidas como *mHealth* (*mobile Health*), essa classe de aplicações oferece uma variedade de

serviços, incluindo telemedicina, lembretes de saúde, educação em saúde, monitoramento de condições crônicas e aconselhamento de saúde mental, utilizando *smartphones*, *tablets* e outros dispositivos móveis [Organization et al. 2015, Kim et al. 2024].

Estudos destacam a importância da motivação do paciente para o autocuidado em saúde, relacionada às ações realizadas para promover e melhorar o bem-estar [Riegel et al. 2021, Woldaregay et al. 2018]. O avanço tecnológico na área da saúde impulsiona essa prática, especialmente com a crescente adoção de dispositivos móveis por pessoas de todas as idades [Machado et al. 2022].

A aplicação *mHealth* IUProst¹ foi desenvolvida com o propósito de fornecer suporte aos pacientes que passam pelo tratamento da incontinência urinária, resultante da cirurgia de retirada da próstata (prostatectomia radical), em virtude do câncer. As funcionalidades do aplicativo englobam uma variedade de recursos essenciais para o acompanhamento e gerenciamento da condição de saúde do paciente. Isso inclui orientações detalhadas sobre o tratamento da incontinência urinária, prescrição de exercícios diários específicos, acompanhamento do engajamento do paciente em seu regime terapêutico, e o monitoramento contínuo da evolução do quadro clínico do paciente. Ao oferecer essas funcionalidades, o IUProst busca a participação ativa do paciente no processo de autocuidado, visando a melhoria da qualidade de vida e o alcance de melhores resultados na reabilitação da incontinência urinária pós-prostatectomia radical.

Iniciativas de aplicações *mHealth*, em geral, necessitam da efetividade da aplicação enquanto autocuidado. Pesquisas relatam esforços voltados para aplicações *mHealth* com o objetivo de promover efetividade no uso da aplicação, como em [Fan et al. 2024], que investigam a disposição dos usuários em usar aplicativos de saúde móvel, assim como em [dos Anjos et al. 2022], onde os autores apresentam uma investigação de técnicas de aprendizado de máquina e de gamificação que possam prover, efetivamente, recursos para que o paciente se engaje no seu autocuidado.

Os avanços no campo da inteligência computacional, aliados à crescente quantidade de dados gerados por aplicações de saúde em ambientes clínicos, proporcionam o desenvolvimento de sistemas de reconhecimento capazes de identificar padrões ocultos e extrair informações valiosas em bases de dados [Martins et al. 2021]. A Mineração de Dados (MD) é uma metodologia que visa obter dados de qualidade para auxiliar na tomada de decisão, sendo parte do processo de descoberta de conhecimento em bancos de dados, o qual auxilia a transformação de dados em bases de conhecimento (do inglês, *Knowledge Discovery in Databases* - KDD). Em [Gupta and Chandra 2020], a metodologia é descrita como a busca e extração de conhecimento previamente desconhecido a partir de dados armazenados.

A mineração de dados requer uma abordagem padrão que ajude a traduzir os problemas de negócios em tarefas de mineração de dados, além de sugerir as transformações de dados apropriadas para se obter o conhecimento do negócio [Wirth and Hipp 2000]. Nesse sentido, a metodologia CRISP-DM (do inglês, *Cross Industry Standard Process for Data Mining*) é um modelo de processo independente do domínio para aplicação em projetos de MD [Schröer et al. 2021]. As tarefas previstas para cada fase do processo apresentam resultados que colaboram para a melhoria e evolução das aplicações.

¹<https://www.iuprosth.com.br/>

O objetivo deste artigo é aplicar a metodologia CRISP-DM para identificar padrões, tendências e percepções nos dados do IUProst², visando levantar demandas para subsidiar os próximos ciclos de evolução da aplicação, em especial quanto à sua efetividade no autocuidado. Os resultados obtidos indicam que a utilização dos dados do IUProst em um processo de mineração de dados padrão como o CRISP-DM, colabora nesse sentido e na elucidação de problemas e de melhorias na aplicação. Os resultados são animadores, pois apontam, para além do IUProst, em direção a um modelo de MD potencialmente válido e replicável para a classe de aplicações *mHealth*.

As próximas seções do artigo estão estruturadas da seguinte forma: a Seção 2 apresenta o IUProst, detalha a metodologia CRISP-DM, e discute trabalhos que utilizam esta metodologia em aplicações *mHealth*; a Seção 3, por sua vez, descreve o uso do CRISP-DM na análise de dados do IUProst; as discussões dos resultados são apresentados na Seção 4; e, por fim, na Seção 5 apresenta-se a conclusão e trabalhos futuros.

2. Contextualização

Nesta seção, são apresentadas as principais características e funcionalidades do IUProst. A metodologia CRISP-DM e suas fases serão detalhadas, seguidas da descrição de estudos relacionados, com destaque para o seu uso em aplicações de saúde.

2.1. Aplicativo *mHealth* IUProst

O IUProst é uma aplicação *mhealth* cujo objetivo é auxiliar o tratamento da incontinência urinária, comorbidade identificada em homens que realizaram a cirurgia de prostatectomia radical (retirada da próstata). O aplicativo faz parte de um projeto desenvolvido em parceria pela Escola de Enfermagem da Universidade Federal de Minas Gerais (UFMG) e o Laboratório de Informática e Saúde do Instituto de Informática da Universidade Federal de Goiás (LabIS-INF-UFG). Os resultados das pesquisas, realizadas de forma interdisciplinar, já resultaram em dissertações de mestrado, teses de doutorado, além de publicações nacionais e internacionais.

O câncer de próstata, conforme as estimativas apresentadas por [de Oliveira Santos et al. 2023], é reconhecido como um dos mais prevalentes, com a expectativa de 71 mil novos casos no período de 2023 a 2025. O estudo relata que este tipo de câncer é a forma mais comum entre os homens, destacando sua significância como um sério problema de saúde pública em escala global. Em decorrência da cirurgia de retirada da próstata, homens podem apresentar o distúrbio da incontinência urinária (IU). Abordagens terapêuticas baseadas em terapias comportamentais são recomendadas pela *International Continence Society* (ICS) como tratamento inicial da IU [Izidoro et al. 2022]. O tratamento da IU por meio da terapia comportamental visa auxiliar o paciente na modificação do estilo de vida e no treinamento muscular do assoalho pélvico. Essas ações, requerem que o paciente esteja consciente de sua condição e envolvido com o tratamento.

As funcionalidades do aplicativo, divididas em três categorias (usuários, tratamento cognitivo comportamental e questionários) são apresentadas na Figura 1. As funcionalidades relacionadas a Usuários englobam as rotinas comuns de cadastro,

²Aprovação pelo Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais (COEP-UFMG), parecer n.º r n.º 4.864.981, CAAE: 41736921.5.0000.5149

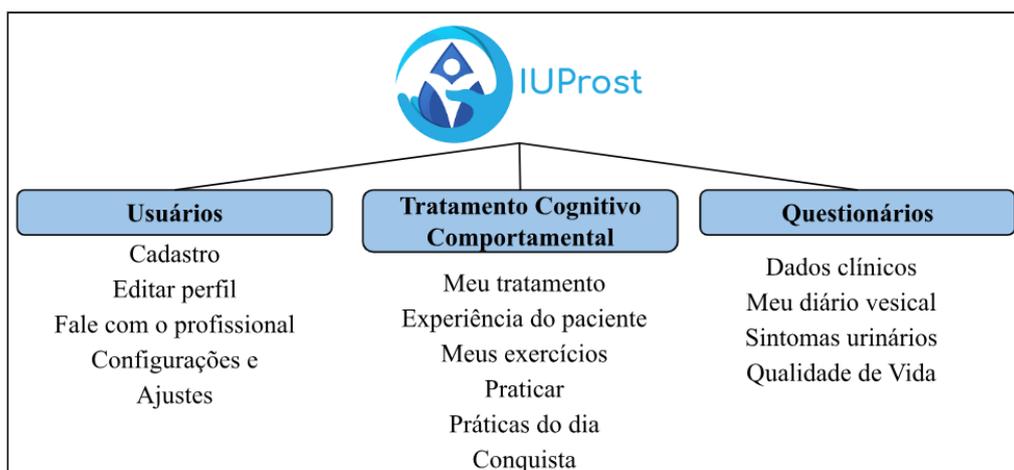


Figura 1. Funcionalidades do IUProst.

configurações e contato com o profissional de saúde. Em Tratamento Cognitivo Comportamental estão agrupadas as funcionalidades referentes a: informações sobre o tratamento; relatos de pacientes que já realizaram o tratamento; visualização do desempenho semanal durante o tratamento; orientações para execução dos exercícios; visualização do desempenho com percentual cumprido por período do dia; e as conquistas adquiridas ao longo da semana. Os questionários possibilitam o relato de informações clínicas dos pacientes durante o tratamento.



Figura 2. Telas do IUProst.

A Figura 2 apresenta quatro telas do aplicativo IUProst:

- Tela 1: funcionalidades disponíveis e seus ícones;
- Tela 2: orientações para o usuário realizar a prática dos exercícios, com a descrição da etapa em texto e em vídeo, além de orientações sonoras emitidas durante a prática;
- Tela 3: evolução das atividades realizadas pelo usuário, na forma de um gráfico; e

- Tela 4: questionários que captam informações sobre os líquidos ingeridos e o volume, e condições de perda de urina.

O desenvolvimento do IUProst utilizou o *framework React Native (front-end)*, *Node.js (back-end)*, e para o sistema de gerenciamento de bancos de dados foi utilizado o *MySQL*. O IUProst, lançado na *Google Play Store*³ no ano de 2022, tem sido amplamente utilizado, gerando um conjunto significativo de dados dos usuários cadastrados.

2.2. CRISP-DM

O CRISP-DM estabelece um padrão para a condução de projetos de mineração de dados (MD). Seu objetivo é produzir conhecimento relevante para dar suporte a processos de tomada de decisão. O modelo do processo serve como base para a elaboração de um modelo especializado que detalha as etapas a serem seguidas, oferecendo orientações práticas para cada uma delas. Essa metodologia foi selecionada devido a algumas de suas características: padronização dos processos, que auxilia na replicação do método; métricas e métodos de avaliação claros; estrutura bem definida do que investigar e analisar; e o potencial de aplicação de modelos de mineração de dados em cenários reais [Schröer et al. 2021].

Seis fases interativas compõem o modelo: (i) compreensão do negócio; (ii) compreensão dos dados; (iii) preparação dos dados; (iv) modelagem; (v) avaliação; e (vi) implantação [Wirth and Hipp 2000]. A Figura 3 apresenta as interações entre as fases e a Tabela 1 identifica os objetivos e saídas de cada fase. As interações entre as fases ocorrem durante a execução do projeto.

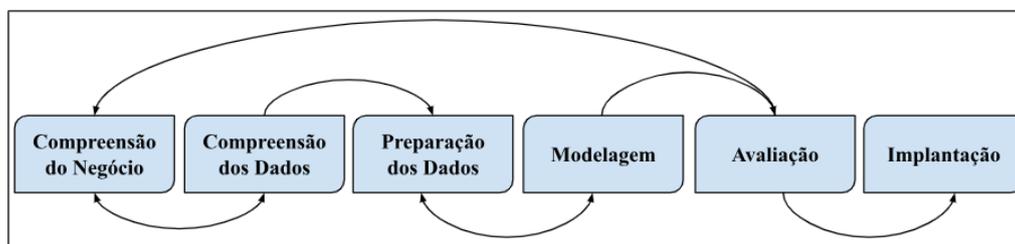


Figura 3. Fases interativas do CRISP-DM. Adaptado de [Plotnikova et al. 2022].

2.3. Estudos Relacionados

Há uma variedade de estudos que relatam experiências de uso do CRISP-DM, em domínios como a educação [Ramos et al. 2020], finanças [Plotnikova et al. 2022], indústria [Maataoui et al. 2023], entre outros. Nesta seção, são apresentados estudos relacionados ao emprego da metodologia CRISP-DM em projetos na área da saúde. Os estudos foram selecionados com base na sua relevância para os objetivos desta pesquisa, utilizando palavras-chave pertinentes em bases de dados científicas. Ao final, as abordagens dos estudos investigados é comparada com o uso proposto nesta pesquisa.

Em [Martins et al. 2021], os autores realizaram um estudo que tem como foco a aplicação de técnicas de MD a dados clínicos coletados durante exames médicos realizados em indivíduos para detectar a presença de doença cardiovascular. Para isso, foi

³https://play.google.com/store/apps/details?id=com.devfilsk.iuprost&hl=pt_BR&gl=US

Tabela 1. Descrições e saídas de cada fase da metodologia CRISP-DM. Adaptado de [Schröer et al. 2021].

Fase	Descrição	Abordagem
Compreensão do Negócio	Avaliação do negócio, recursos disponíveis e necessários. Os objetivos do projeto e da mineração de dados precisam ser definidos.	Documento de projeto de dados; objetivos da mineração de dados.
Compreensão dos Dados	Identificação da fonte de dados para coleta, exploração e descrição. A qualidade dos dados deve ser verificada. Descrição de dados feita por meio de análise estatística e determinação dos atributos e seus agrupamentos.	Descrição da fonte de dados e processo de coleta; descrição estrutural (modelo de dados, dados de exemplo); estatísticas descritivas.
Preparação dos Dados	Seleção dos dados pela definição de critérios de inclusão e exclusão. Limpeza de dados e construção de atributos derivados, se necessário. A estruturação do conjunto de dados depende do modelo a ser implementado.	Descrição dos dados de entrada e saída; métodos e abordagens (transformação, seleção, limpeza).
Modelagem	Seleção da técnica de modelagem; construção do caso de teste e do modelo. Escolha da técnica de mineração de dados depende do modelo e do negócio. Para avaliar o modelo é apropriado avaliar o modelo em relação aos critérios de avaliação.	Abordagem de modelagem; tecnologia utilizada para execução do modelo; construção de conjuntos de teste e treinamento.
Avaliação	Verificação dos resultados em relação aos objetivos de negócio definidos. Os resultados devem ser interpretados e outras ações devem ser definidas. O processo deve ser revisto de forma geral.	Definição de métricas; visualização do modelo e métricas.
Implantação	Relatório final ou um componente de software.	Descrição das implementações no projeto ou no guia do usuário.

aplicada a metodologia CRISP-DM utilizando-se uma base de dados pública. Classificadores foram aplicados para a detecção de doença cardiovascular. Esse estudo apresentou resultados promissores utilizando algoritmos de árvore de decisão para o diagnóstico de doença cardiovascular.

No estudo de [Peixoto et al. 2022], os autores dão continuidade a um estudo cujo objetivo é diversificar a aplicação de técnicas de MD para se identificar a ocorrência de diabetes em pacientes. O objetivo do estudo é prever diagnosticamente se um paciente possui ou não diabetes, considerando características como insulina, nível, concentração de glicose plasmática, pressão arterial, espessura da pele, entre outras. O estudo utiliza um conjunto de dados públicos, denominado *The Pima Indians Diabetes Dataset*, e aplica e analisa seis modelos de aprendizado de máquina em diferentes abordagens dos dados. O conjunto de dados foi usado para testar sete algoritmos de classificação diferentes.

No estudo de [Oliveira et al. 2022], a metodologia CRISP-DM foi utilizada pelos autores para conduzir e desenvolver um modelo preditivo capaz de identificar infecções por COVID-19 conforme a gravidade da doença. No estudo, dados sobre o estado evolutivo dos pacientes em atendimento domiciliar, bem como os resultados do seu teste de COVID-19, foram utilizados para treinar diferentes algoritmos de aprendizado de máquina. Os autores observaram que dentre os algoritmos analisados, a árvore de decisão apresentou maior precisão, entre os critérios de validação. Como trabalho futuro, é descrito no estudo a implementação de um sistema de apoio a decisão com base no modelo preditivo construído.

Trabalhos que exploram o uso da metodologia CRISP-DM apresentam um problema de MD e aplicam algoritmos de aprendizado de máquina para treinar algoritmos e construir modelos para detecção de padrões relacionados aos problemas investigados. Na abordagem proposta neste estudo, busca-se ampliar o uso da metodologia para que, a partir do entendimento do negócio e dos dados, seja possível identificar um conjunto de problemas de mineração de dados, extrair os dados do banco de dados do IUProst, aplicar

e validar algoritmos para subsidiar as melhorias do software em termos de interação, engajamento e recomendação de boas práticas no uso do software para obtenção de melhores resultados no tratamento.

As descobertas realizadas pelo uso da metodologia CRISP-DM, darão subsídio às implementações necessárias no IUProst, que serão realizadas na fase da implantação. A previsão da implantação se apresenta como outro ponto forte deste projeto, visto que os estudos investigados não detalham essa fase da metodologia.

3. Uso de CRISP-DM na análise de dados do IUProst

Nesta seção, são descritas as etapas realizadas na análise de dados e os resultados alcançados em cada uma delas. A metodologia CRISP-DM foi utilizada neste estudo para auxiliar no planejamento e gerenciamento das tarefas, visando produzir resultados que subsidiem a evolução do IUProst. Os próximos tópicos detalham as fases da metodologia executadas e os resultados produzidos.

3.1. Resultados da Fase de Compreensão do Negócio

O esforço desta etapa se concentra em entender as intenções e os requisitos do projeto, do ponto de vista do negócio, para converter esse conhecimento nos objetivos da MD.

O tratamento cognitivo comportamental é composto por um ciclo de 8 semanas (etapas) em que o paciente realiza diariamente exercícios para o treinamento da musculatura do assoalho pélvico. Os protocolos preveem a execução de exercícios específicos em cada etapa do tratamento, nos períodos matutino, vespertino e noturno. São apresentados textos explicativos sobre a IU e o tratamento, vídeos explicativos e orientações sonoras no momento da execução dos exercícios, além de depoimentos de pacientes que tiveram IU e realizaram o tratamento.

O aplicativo registra e apresenta o desempenho do paciente em cada dia e semana de tratamento, apresentando mensagens com orientações e, também, para motivá-lo. Durante as etapas são disponibilizados questionários para registro de ingestão de líquidos, perda de urina, uso de absorventes e fraldas, dentre outros. Um documento denominado projeto de dados foi desenvolvido contendo as descrições importantes do IUProst e as ações de cada fase da metodologia.

Um dos principais desafios identificados no IUProst é a entrega aos usuários de experiências de interações agradáveis, que promovam o uso do IUProst e, conseqüentemente, a efetividade no tratamento da IU. Desta forma, foram identificados os seguintes objetivos da MD:

- identificar as características mais relevantes entre os usuários que finalizaram o ciclo de exercícios propostos no tratamento;
- agrupar usuários de acordo com perfis de uso da aplicação; e
- construir um modelo que, quando aplicados aos dados de novos usuários, realize a recomendação de ações no IUProst, de maneira personalizada.

Assim, o objetivo é desenvolver uma solução que se adapte às experiências dos usuários para melhorar a aplicação com retornos personalizados, promovendo a navegação adaptada ao perfil do usuário. Além disso, deve-se possibilitar ao profissional de saúde o acompanhamento das ações realizadas pelo usuário.

3.2. Resultados da Fase de Compreensão dos Dados

Os estudos apresentados anteriormente, como [Martins et al. 2021] e [Peixoto et al. 2022], utilizam conjuntos de dados públicos. Para este estudo, foi necessário compreender os dados a partir da documentação e do modelo de entidade e relacionamentos do IUProst. Visto que, para o IUProst, este é o primeiro estudo na vertente de análise de dados, esta fase entrega subsídios para a fase de preparação dos dados, além de documentação de suporte para pesquisas futuras.

Para apoiar o entendimento dos dados, o dicionário de dados do projeto foi utilizado, assim como reuniões com a equipe para entendimento dos fluxos de dados do aplicativo. Foi realizada a anonimização dos dados para a construção de consultas em SQL (*Structured Query Language*), que selecionaram as informações relevantes para construção dos conjuntos de dados para as análises.

Os atributos identificados como relevantes para compor os conjuntos de dados foram agrupados em categorias: dados do paciente, dados do tratamento e dados dos questionários. A Tabela 2, apresenta a descrição de cada atributo identificado até o momento.

Tabela 2. Características dos atributos.

Atributo	Descrição	Tipo
id	Identificação única do paciente ¹	dados paciente
idade	Idade do paciente ¹	dados paciente
cirurgia	realizou ou não cirurgia de próstata ²	dados paciente
sonda	usa sonda ²	dados paciente
tratamento	já realizou tratamento de incontinência ² urinária	dados paciente
comorbidade	possui outro tipo de comorbidade ²	dados paciente
peso	peso do paciente ³	dados paciente
altura	altura do paciente ³	dados paciente
IMC	IMC do paciente ³	dados paciente
execucaoexercicios	Quantidade de dias executando os exercícios do tratamento ¹	dados tratamento
etapa	qual etapa do tratamento o paciente se encontra ¹	dados tratamento
ciclo	qual ciclo de execução dos exercícios, o paciente se encontra ¹	dados tratamento
perdaurainai	perda de urina (início do tratamento) ²	dados questionários
perdaurinafim	perda de urina (fim do tratamento) ²	dados questionários

Tipos de Dados dos atributos: ¹{Valor numérico inteiro.} ²{1 - Sim;0 - Não} ³{Valor numérico real.}

Para a melhor compreensão de cada atributo, na Tabela 3 é apresentada a quantidade de valores ausentes, bem como os valores mínimo e máximo, média e desvio padrão para as principais propriedades do tipo de dados paciente. Os valores relativos ao desvio padrão identificados para idade, peso, altura e IMC sugerem dispersões moderadas em torno da média dos grupos.

Tabela 3. Estatística descritiva de atributos

Atributo	Valores Ausentes	Min	Max	Média	Desvio Padrão
idade	13	21	91	54	16,72
peso	0	50	129	78,88	13,29
altura	0	1,50	1,95	1,71	0,06
IMC	0	17,30	43,10	26,29	3,95

Ao realizar consultas, observou-se que do mês de outubro de 2022 ao mês de

março de 2024, o aplicativo IUProst atingiu aproximadamente 1000 usuários cadastrados e mais de 24 mil exercícios executados.

Entre os cadastros realizados, 889 usuários passaram pela etapa de preenchimento do questionário de dados clínicos e de informações de saúde. Dentre estes, 547 usuários (61,5%) relataram terem sido submetidos à cirurgia de retirada da próstata. Este percentual representa, de fato, usuários que compõem o público-alvo e representam pacientes potenciais para o tratamento fornecido pelo IUProst.

Além da identificação do público-alvo, também foi realizada a análise dos usuários que se envolveram com o tratamento proposto pela aplicação, compondo aproximadamente 8% do total do público-alvo.

4. Discussão dos Resultados

Nesta Seção são discutidos os resultados alcançados nesta pesquisa, que relatam a experiência da execução de duas fases da metodologia CRISP-DM.

O primeiro ponto de destaque é o suporte da metodologia para o entendimento do negócio, da ferramenta desenvolvida e dos dados gerados por seu uso. A análise ampliada das etapas iniciais indica, positivamente, que o uso da metodologia dá suporte para a identificação de lacunas a serem exploradas. Isso contribui para a evolução da aplicação *mHealth* de forma embasada e direcionada a fim de aprimorar a solução do negócio (o IUProst), para que esse possa colaborar de forma mais eficaz no tratamento da IU e, desta forma, colaborar com a melhora da qualidade de vida dos homens com essa comorbidade.

Os resultados apresentados na Seção 3.1, que representam a execução da primeira fase do CRISP-DM, demonstram a importância do estudo do negócio, para serem definidos novos objetivos para o projeto. Durante esta etapa, foram enfrentadas dificuldades de ausência de documentação sobre o aplicativo, fato superado pela realização de entrevistas com integrantes do projeto e participação em reuniões. Foi identificada também a necessidade de documentar a investigação realizada e, desta forma, os objetivos definidos para a MD.

Na Seção 3.2 foram descritas as principais atividades realizadas na segunda fase da metodologia, a qual identificou o conjunto de dados gerados pela utilização do IUProst, por meio da seleção e descrição dos principais atributos que podem colaborar com a investigação. As análises consideram os dados que identificam o público-alvo do IUProst, e o envolvimento destes no tratamento proposto. Os números mostraram uma quantidade reduzida de pacientes que realizaram todo o processo fornecido pelo aplicativo, o tratamento cognitivo comportamental. Esse fato sugere, portanto, uma baixa adesão à terapia sugerida, por conseguinte, um baixo engajamento na alternativa para o autocuidado.

Essas percepções iniciais confirmam a necessidade da análise dos dados empregada para superar os desafios inerentes as aplicações *mHealth* e, de forma específica, colaborar para o alinhamento dos objetivos da MD definidos na primeira fase.

5. Conclusão e trabalhos futuros

A análise de dados do aplicativo IUProst, conduzida nesta pesquisa empregando a metodologia CRISP-DM, visa identificar demandas que orientem a evolução e o aprimora-

mento da aplicação, na direção de ajudar pessoas em tratamento de incontinência urinária, e colaborar, assim, com a melhoria de um problema de saúde pública.

Os resultados iniciais coletados durante a primeira fase do CRISP-DM demonstram a importância do estudo do negócio. Durante esta etapa, foram enfrentadas dificuldades devido à ausência de documentação do aplicativo, que foi superada pela realização de entrevistas com integrantes do projeto e participação em reuniões. Identificou-se também a necessidade de documentar a investigação de uma forma mais ampla e, dessa forma, definir os objetivos para a mineração de dados. As principais atividades realizadas na segunda fase da metodologia incluíram a identificação do conjunto de dados gerados pela utilização do IUProst, com a seleção e descrição dos principais atributos. As análises iniciais revelaram uma baixa participação do público-alvo no tratamento proposto, confirmando a necessidade da análise de dados para superar desafios inerentes às aplicações *mHealth*.

Após a execução das duas fases propostas pela metodologia, é possível confirmar suspeitas iniciais de que são necessários mecanismos que promovam maior participação no tratamento pelo público-alvo deste objeto de estudo. Esse indício reforça a necessidade da continuidade do desenvolvimento do software. Ao final das etapas, com os resultados da aplicação das fases de modelagem e avaliação propostos pelo CRISP-DM, pretende-se utilizar os resultados para a implantação (última fase) de soluções alinhadas com pesquisas realizadas anteriormente no IUProst, que exploraram a gamificação e a avaliação da usabilidade. Também está em discussão, a construção de um sistema de *chat* conversacional baseado no conhecimento descoberto, para aprimorar a interação com o usuário e, buscando uma participação mais efetiva deste no seu autocuidado.

Por fim, a experiência e os resultados alcançados neste esforço são promissores, por vislumbrarem cenários em direção a um modelo de mineração de dados replicável e orientado para aplicações *mHealth* focadas no autocuidado do usuário, considerando que as principais melhorias percebidas passam pelo processo de conhecer o dados gerado pelo IUProst.

Referências

- de Oliveira Santos, M., de Lima, F. C. d. S., Martins, L. F. L., Oliveira, J. F. P., de Almeida, L. M., and de Camargo Cancela, M. (2023). Estimativa de incidência de câncer no Brasil, 2023-2025. *Revista Brasileira de Cancerologia*, 69(1).
- dos Anjos, F. M., Oliveira, L. W., Souza, C. H. R., and Carvalho, S. T. (2022). Aplicações de técnicas de machine learning e gamificação no autocuidado em saúde: Uma revisão sistemática. *Anais Estendidos do XXI Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 1376–1380.
- Fan, S., Jain, R. C., and Kankanhalli, M. S. (2024). A comprehensive picture of factors affecting user willingness to use mobile health applications. *ACM Transactions on Computing for Healthcare*, 5(1):1–31.
- Gupta, M. K. and Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, 12(4):1243–1257.
- Izidoro, L. C. d. R., Mata, L. R. F. d., Azevedo, C., Paula, A. A. P. d., Pereira, M. G., Santos, J. E. M. d., Brasil, V. V., and Oliveira, L. M. d. A. C. (2022). Programa cognitivo-

- comportamental para controle de sintomas do trato urinário inferior pós-prostatectomia radical: ensaio clínico randomizado. *Revista Brasileira de Enfermagem*, 75.
- Kim, G., Hwang, D., Park, J., Kim, H. K., and Hwang, E.-S. (2024). How to design and evaluate mhealth apps? a case study of a mobile personal health record app. *Electronics*, 13(1).
- Maataoui, S., Bencheikh, G., and Bencheikh, G. (2023). predictive maintenance in the industrial sector: a crisp-dm approach for developing accurate machine failure prediction models. In *2023 Fifth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, pages 223–227. IEEE.
- Machado, A. F., Estevam, F. E. B., Izidoro, L. C. d. R., Oliveira, H. M. d., Anjos, F. M. d. S. d., Carvalho, S. T. d., and Mata, L. R. F. d. (2022). Male urinary incontinence and the digital technology: Evaluation of mobile applications available for download. *Cogitare Enfermagem*, 27.
- Martins, B., Ferreira, D., Neto, C., Abelha, A., and Machado, J. (2021). Data mining for cardiovascular disease prediction. *Journal of Medical Systems*, 45:1–8.
- Oliveira, D., Ferreira, D., Abreu, N., Leuschner, P., Abelha, A., and Machado, J. (2022). Prediction of covid-19 diagnosis based on openehr artefacts. *Scientific Reports*, 12(1):12549.
- Organization, W. H. et al. (2015). A practical guide for engaging with mobile operators in mhealth.
- Peixoto, H., Ramos, V., Marques, C., and Machado, J. (2022). Predicting diabetes disease for healthy smart cities. *EAI Endorsed Transactions on Smart Cities*, 6(18):e1–e1.
- Plotnikova, V., Dumas, M., and Milani, F. P. (2022). Applying the crisp-dm data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & knowledge engineering*, 139:102013.
- Ramos, J. L. C., Rodrigues, R. L., Silva, J. C. S., and de Oliveira, P. L. S. (2020). Crisp-dm: uma proposta de adaptação do modelo crisp-dm para mineração de dados educacionais. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1092–1101. SBC.
- Riegel, B., Dunbar, S. B., Fitzsimons, D., Freedland, K. E., Lee, C. S., Middleton, S., Stromberg, A., Vellone, E., Webber, D. E., and Jaarsma, T. (2021). Self-care research: where are we now? where are we going? *International journal of nursing studies*, 116:103402.
- Schröer, C., Kruse, F., and Gómez, J. M. (2021). A systematic literature review on applying crisp-dm process model. *Procedia Computer Science*, 181:526–534.
- Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester.
- Woldaregay, A. Z., Issom, D.-Z., Henriksen, A., Marttila, H., Mikalsen, M., Pfuhl, G., Sato, K., Lovis, C., and Hartvigsen, G. (2018). Motivational factors for user engagement with mhealth apps. *pHealth*, 249:151–157.