

Antonio Francisco Lima de Oliveira Pádua

**Predição do risco de evasão dos Cursos Técnicos a Distância do
IFPI: uma aplicação da metodologia CRISP-DM em dados
educacionais**

**Recife
2017**



Universidade Federal Rural de Pernambuco
Unidade Acadêmica de Educação a Distância e Tecnologia
Pró-Reitoria de Pesquisa e Pós-Graduação
Programa de Pós-Graduação em Tecnologia e Gestão em Educação a Distância

**Predição do risco de evasão dos Cursos Técnicos a Distância do
IFPI: uma aplicação da metodologia CRISP-DM em dados
educacionais**

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologia e Gestão em Educação a Distância como exigência parcial à obtenção do título de Mestre em Tecnologia e Gestão em Educação a Distância.

**Linha de Pesquisa: Ferramentas Tecnológicas
para Educação a Distância**

Orientador: Prof^o. Dr. Domingos Sávio Salazar

**Recife
2017**

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

P125p Pádua, Antonio Francisco Lima de Oliveira
Predição do risco de evasão dos cursos técnicos a distância do
IFPI: uma aplicação da metodologia CRISP-DM em dados
educacionais / Antonio Francisco Lima de Oliveira Pádua. – 2017.
73 f. : il.

Orientador: Domingos Sávio Salazar.
Dissertação (Mestrado) – Universidade Federal Rural de
Pernambuco, Programa de Pós-Graduação em Tecnologia e Gestão
em Educação a Distância, Recife, BR-PE, 2017.
Inclui referências e apêndice(s).

1. CRISP-DM 2. Evasão 3. Ensino a distância I. Salazar, Domingos
Sávio, orient. II. Título

CDD 371.394422

Universidade Federal Rural de Pernambuco
Unidade Acadêmica de Educação a Distância e Tecnologia
Pró-Reitoria de Pesquisa e Pós-Graduação
Programa de Pós-Graduação em Tecnologia e Gestão em Educação a Distância

**Predição do risco de evasão dos Cursos Técnicos a Distância do
IFPI: uma aplicação da metodologia CRISP-DM em dados
educacionais**

Antonio Francisco Lima de Oliveira Pádua

Dissertação julgada adequada para obtenção do título de Mestre em Tecnologia e Gestão em Educação a Distância, defendida e aprovada por unanimidade em 23/08/2017 pela Banca Examinadora.

Orientador:

Prof. Dr. Domingos Sávio Salazar
Programa de Pós-Graduação em Tecnologia e Gestão em Educação a Distância -
UFRPE

Banca Examinadora:

Prof^a. Dr^a. Juliana Regueira Basto Diniz
Membro Interno – Programa de Pós-Graduação em Tecnologia e Gestão em
Educação a Distância - UFRPE

Prof. Dr. Rodrigo Nonamor Pereira Mariano
Membro Interno – Programa de Pós-Graduação em Tecnologia e Gestão em
Educação a Distância - UFRPE

Prof. Dr. Jorge da Silva Correia Neto
Membro Externo – Programa de Pós-Graduação em Informática Aplicada – UFRPE

Dedico este trabalho à minha avó paterna
Maria dos Anjos Fernandes (*in memoriam*) e
à minha primeira professora Maria Madalena
da Silva.

AGRADECIMENTOS

Agradeço aqui a todos que de alguma forma me apoiaram e incentivaram para alcançar mais um degrau do conhecimento.

A Deus pelas infinitas graças alcançadas.

À família Cavalcanti, em especial ao meu amigo e irmão de coração, Thiago Santos Cavalcanti, que sempre me acolheu em seu lar, antes e durante o período do mestrado.

Ao meu orientador Prof^o Dr. Domingos Sávio Salazar, pela paciência e exemplo de inteligência de um verdadeiro Mestre a ser seguido.

À minha afilhada Andréa Larisse Meneses Rodrigues, que sempre esteve disponível para me ajudar na coleta de dados da minha pesquisa e na organização da minha casa.

A todos do PPGTEG, coordenação, professores e administrativo.

À minha família, sem distinção, pais, irmãos, cunhada e sobrinhos, pelo incentivo em minhas decisões nos estudos e em minha vida. Com o apoio de todos a caminhada fica menos difícil.

Aos meus tios maternos Gilvan Lima de Oliveira e Ivan Lima de Oliveira, pessoas nas quais me espelhei por serem exemplos de dedicação aos estudos.

Aos amigos que conquistei dentro do IFPI: Luciana Pereira, Rosimeyre Vieira, Leidiana Lima, Ana Cláudia Neri, Paulo Roberto Silva, Cristiana Galeno, Glairton, Raquele e Ruthelle. Em especial a Joselma Lavôr, por ser exemplo e incentivadora desde a minha inscrição para a seleção do Mestrado. Amada. Quando eu crescer quero ser como você!

Às minhas amigas do mestrado: Fabiana, Inara, Úrsula, Raqueline, Lilian, Edna Janete e Claudete. Em destaque à minha amiga Cibele Resende, pela companhia em todo o percurso do curso e por ter me dado a oportunidade de conhecer sua mãe Júlia. Com vocês a Europa foi e é logo ali. Acredito que o mestrado foi apenas uma desculpa que o destino usou para que nossas almas se encontrassem.

O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo. Mesmo não se atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas admiráveis.

José de Alencar

RESUMO

Este estudo busca compreender a contribuição da descoberta de conhecimento em bases de dados educacionais aplicando a metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM) no problema da evasão escolar de cursos na modalidade a distância. Assim, teve como objetivo geral adaptar um processo de mineração de dados para estimar o risco de evasão de alunos de Cursos Técnicos de Nível Médio, na modalidade a distância, do Instituto Federal do Piauí (IFPI), sob as etapas da metodologia aludida. Para tanto, foi necessário, entender a descoberta de conhecimento sobre dados de grande volume e a metodologia CRISP-DM como paradigma de mineração de dados educacionais; aplicar a metodologia CRISP-DM para identificar as características comuns de alunos que evadem dos Cursos técnicos a distância do IFPI, para intervenção preventiva da coordenação gestora; e avaliar o risco da evasão de alunos de cursos técnicos na modalidade EAD do IFPI, a partir de dados acadêmicos de alunos ingressantes no ano 2013. A pesquisa possui abordagem quantitativa, documental, descritiva e aplicada. Utilizamos as cinco primeiras fases da metodologia CRISP-DM em dados coletados dos questionários socioeconômicos, e situação final dos alunos no que tange a classificação de evadido ou não evadido após o primeiro semestre do Curso Técnico de Nível Médio em Segurança do Trabalho, na modalidade a distância, do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, *Campus* Teresina Zona Sul. A coleta de dados foi realizada através do Google formulário e integrados utilizando o *software* Excel 2010. Para identificação das variáveis preditoras que explicam a variável alvo, evasão, utilizamos a regressão logística através da função Mínimos Quadrados Ordinários (MQO) do *software* *Greftl* 2016c. Em face deste estudo, é possível dizer que o uso de técnicas de mineração de dados seguindo a metodologia CRISP-DM em base proveniente de cursos a distância também pode contribuir para cursos em quaisquer modalidades no que concerne a busca por estratégias contra a evasão escolar.

Palavras-chave: CRISP-DM. Evasão. Educação a Distância.

ABSTRACT

This study aims to understand the contribution of knowledge discovery in educational databases through the application of the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology to the problem of school dropout from distance learning courses. Thus, the main goal was to adapt a data mining process to estimate the evasion risk of students from Secondary Technical Level Courses, in the Distance Learning modality, on the Federal Institute of Piauí (IFPI), under the aforementioned methodology. To do so, it was necessary to understand the knowledge discovery over high volume data and the CRISP-DM methodology as a paradigm of educational data mining; To apply the CRISP-DM methodology to identify the common characteristics of students who drop out of the technical distance courses of the IFPI, in order to allow the management coordination a chance to proceed with preventive intervention; And to evaluate the risk of evasion of students from technical courses in the Distance Learning modality on the IFPI, based on academic data of incoming students from the year 2013. The research has quantitative, documental, descriptive and applied approaches. We used the first five phases of the CRISP-DM methodology in data collected from the socioeconomic questionnaires, and the final situation of the students regarding the classification of evaded or not evaded after the first semester of the Distance Learning Technical Course in Work Safety of the Federal Institute of Education, Science and Technology of Piauí, Campus Teresina Zona Sul. Data collection was done through Google form and the data integration was done using the Excel 2010 software. In order to identify the predictor variables that explain the target variable, school dropout, we used logistic regression through the Ordinary Least Squares (OLS) function of Gretl 2016c software. In light of this study, it is possible to say that the use of data mining techniques following the CRISP-DM methodology based on distance learning courses can also contribute to courses in any modalities regarding the search for strategies against school dropout.

Keywords: CRISP-DM. School Dropout. Distance Learning.

LISTA DE FIGURAS

Figura 1 - Descoberta de Conhecimento em Bases de Dados.....	29
Figura 2 - Pirâmide da Informação	30
Figura 3 – Metodologia CRISP-DM.....	34
Figura 4 - Preferência pela Metodologia CRISP-DM.....	39
Figura 5 - Metodologia CRISP-DM aplicada a EAD do IFPI.....	40

LISTA DE TABELAS

Tabela 1 - Dados relativos à educação no relatório do PNUD 2012	16
---	----

LISTA DE QUADROS

Quadro 1 - Causas da evasão escolar por dimensão.....	23
Quadro 2 - Tarefas do Entendimento do Negócio	41
Quadro 3 - Tarefas do Entendimento dos Dados	42
Quadro 4 - Tarefas da Preparação dos Dados.....	43
Quadro 5 - Dicionário Preliminar de Variáveis.....	43
Quadro 6 - Dicionário Final de Variáveis	45
Quadro 7 - Tarefas da Modelagem	46
Quadro 8 - Passo a passo da mineração de dados	56

LISTA DE EQUAÇÕES

Equação 1 - Regressão logística.....	32
Equação 2- Função logística	33

LISTA DE GRÁFICOS

Gráfico 1 - Distribuição de alunos por faixa de score	51
---	----

LISTA DE SIGLAS

AVA	Ambiente Virtual de Aprendizagem
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
DCBD	Descoberta do Conhecimento em Base de Dados
DM	<i>Data Mining</i>
EAD	Educação a Distância
EDM	<i>Educational Data Mining</i>
ENEM	Exame Nacional do Ensino Médio
IDH	Índice de Desenvolvimento Humano
IFPI	Instituto Federal de Educação, Ciência e Tecnologia do Piauí
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
KDD	<i>Knowledge Discovery in Databases</i>
MD	Mineração de Dados
MDE	Mineração de Dados Educacionais
MEC	Ministério da Educação
MQO	Mínimos Quadrados Ordinários
PNUD	Programa das Nações Unidas para o Desenvolvimento
SETEC	Secretaria de Educação Profissional e Tecnológica
UFRPE	Universidade Federal Rural de Pernambuco

SUMÁRIO

1	INTRODUÇÃO	15
1.1	JUSTIFICATIVA	17
1.2	OBJETIVOS DA PESQUISA	19
1.3	ESTRUTURA DA DISSERTAÇÃO	20
2	REFERENCIAL TEÓRICO	21
2.1	EVASÃO ESCOLAR: PERCEPÇÕES E DIMENSÕES	21
2.2	MINERAÇÃO DE DADOS	26
2.3	MINERAÇÃO DE DADOS EDUCACIONAIS	31
2.4	CRISP-DM.....	33
3	METODOLOGIA DA PESQUISA	37
3.1	CARACTERIZAÇÃO E CONTEXTUALIZAÇÃO DA PESQUISA	37
3.2	APLICAÇÃO DA METODOLOGIA CRISP-DM.....	39
3.2.1	Entendimento do Negócio	40
3.2.2	Entendimento dos Dados	41
3.2.3	Preparação dos Dados	42
3.2.4	Modelagem	46
3.2.5	Avaliação	47
3.2.6	Implantação	48
4	ANÁLISE E DISCUSSÕES DOS RESULTADOS	49
5	CONSIDERAÇÕES FINAIS	53
	REFERÊNCIAS	58
	APÊNDICE A – FOTOS DA COLETA DOS DADOS	63
	APÊNDICE B – TERMO DE AUTORIZAÇÃO DO USO DE IMAGEM	64
	APÊNDICE C – FORMULÁRIO ONLINE DA COLETA DOS DADOS	65
	APÊNDICE D – RESUMO DOS DADOS COLETADOS	66
	APÊNDICE E – GERENCIAMENTO DAS LEITURAS	68
	APÊNDICE F – RESUMO DO GERENCIAMENTO DAS LEITURAS	69
	APÊNDICE G – VARIÁVEIS FORTEMENTE PREDITIVAS	70
	APÊNDICE H – PRODUTO DA PESQUISA	71

1 INTRODUÇÃO

A evasão escolar é tema de debates e reflexões na educação brasileira, ocupando um espaço de relevância no cenário das políticas públicas, no qual as dificuldades de acesso e permanência do aluno na Educação Básica apresentam marcas significativas da fragilidade do nosso sistema educacional, por persistir ao longo da história da educação brasileira.

Nesse sentido, visualizamos que por ser um fenômeno complexo, a evasão pode desencadear consequências indesejáveis à vida acadêmica e profissional dos alunos, bem como produzir a médio e longo prazo prejuízos sociais e econômicos.

Segundo Johann (2012), o que está acontecendo é que as instituições estão investindo nas salas de aula prevendo a sua lotação, todavia, deparando-se com prejuízo e ociosidade já no primeiro semestre, considerando nesse contexto que se avolumam as questões que desafiam a permanência e o êxito dos estudantes.

Corroborando nesse entendimento Silva Filho, Lobo e Hipólito (2009), afirmam que independentemente da instituição ser pública ou privada, a evasão escolar em ambos os casos, implica em desperdícios sociais, acadêmicos e econômicos. Nesse sentido, percebemos que o ato de evadir-se não implica apenas “na perda do aluno”, mas também, de um futuro profissional com grandes oportunidades de qualificação para atender às demandas do mercado de trabalho, bem como no desperdício de todas as potencialidades que as instituições públicas ou privadas fazem para proporcionar um ensino de qualidade com motivação para que os estudantes permaneçam na escola.

Nessa perspectiva, cresce continuamente a necessidade de se vislumbrar no campo da concretude, a tríade: acesso, permanência e êxito, pois como Castro e Tiezzi (2005) mencionam, a educação deve ser permanente na vida das pessoas, viabilizando frequentes retornos a fim de desenvolver o raciocínio, o pensamento crítico e a contextualização dos conhecimentos adquiridos.

Justifica-se o tema evasão está numa agenda propositiva, haja vista, por exemplo, no que tange à educação mundial, que o Programa das Nações Unidas para o Desenvolvimento (PNUD) emitiu o Relatório de Desenvolvimento 2012, elencando a taxa de evasão escolar de 100 países. De acordo com essas taxas, algumas conclusões foram obtidas e uma delas mostrou que, no Brasil, a cada quatro alunos que iniciam o ensino fundamental, um abandona a escola antes de completar a última

série deste mesmo nível quando comparado com outros países (PORTAL TODOS PELA EDUCAÇÃO, 2013).

Tabela 1 - Dados relativos à educação no relatório do PNUD 2012

País	Posição no ranking	IDH	População alfabetizada	População com pelo menos ensino médio completo	Taxa de evasão escolar
Noruega	1º	0,955	100%	95,2%	0,5%
Austrália	2º	0,938	100%	92,2%	Não informada
Estados Unidos	3º	0,937	100%	94,5%	6,9%
Holanda	4º	0,921	100%	88,9%	Não informada
Alemanha	5º	0,920	100%	96,5%	4,4%
Chile	40º	0,819	98,6%	74%	2,6%
Argentina	45º	0,811	97,8%	56%	6,2%
Uruguai	51º	0,792	98,1%	49,8%	4,8%
México	61º	0,775	93,1%	53,9%	6%
Brasil	85º	0,730	90,3%	49,5%	24,3%

Fonte: PNUD (2012).

Conforme a **Tabela 1**, podemos perceber que o percentual da evasão escolar compromete outras taxas como o Índice de Desenvolvimento Humano (IDH) e o percentual da população que possui pelo menos o Ensino Médio completo, por observarmos uma relação de proporcionalidade inversa quando comparadas com o tema do nosso trabalho.

Além disso, o índice (24,3%) de evasão escolar no Brasil mostra que a educação não está sendo plena no que se refere ao alcance de todos os cidadãos, comprometendo assim, a obtenção de outros níveis de escolaridade como a graduação e a pós-graduação.

Diante deste fato, em 2013, o Ministério da Educação (MEC) criou um grupo de trabalho formado por representantes da Secretaria de Educação Profissional e Tecnológica (SETEC) e da Rede Federal de Educação Profissional, Científica e Tecnológica com o objetivo de emitir um relatório dos índices de evasão escolar e elaborar um manual de orientação contra a evasão (PORTAL BRASIL, 2013).

Em geral, organizar relatórios para analisá-los com o objetivo de extrair conclusões é um processo que requer tempo, principalmente quando há um grande volume de dados a serem analisados e estes estão em formato rudimentar e distintos. Ademais, esta ação pode tornar-se mais difícil quando depende exclusivamente do esforço humano.

Nessa lógica, entendemos que é importante o desenvolvimento e uso de ferramentas tecnológicas e procedimentos que permitam extrair conhecimentos necessários para a identificação precoce de alunos que apresentam tendências a abandonar os estudos, pois a evasão é um fenômeno silencioso que não escolhe a modalidade escolar, ela apenas usa o seu poder de invisibilidade para desestruturar o ensino, e alterar significativamente a trajetória acadêmica e profissional de não poucos.

Refletindo nesse âmbito percebemos que, ao utilizarmos a Mineração de Dados (MD) em registros acadêmicos de estudantes, é possível identificarmos padrões e correlações existentes que, em consequência podem promover ações de gestão mais conscientes e consistentes nas instituições educativas.

Nossa defesa pela Mineração de Dados Educacionais (MDE) converge nessa direção, pois pode ser vista como uma das várias estratégias utilizadas para prevenir e reduzir a evasão, independentemente da escola ser pública ou privada, ou da modalidade de ensino adotada na oferta de seus cursos.

1.1 JUSTIFICATIVA

A evasão escolar tem uma natureza complexa, pois a escolha de sair da escola é apenas a atitude final de um processo que se manifesta, visivelmente ou não, ao longo da vida escolar do aluno.

Abandonar ou permanecer na escola é uma escolha fortemente condicionada não só por características individuais, sociais e familiares, mas também por características do sistema escolar.

Uma das principais dificuldades das instituições de ensino é a identificação do risco de evasão, visando manter equivalente o percentual de alunos matriculados com a quantidade de egressos. Isto decorre da incapacidade em estimar-se adequadamente a probabilidade dos discentes concluírem ou não um curso dos mais diversos níveis.

Com isso, entendemos que as instituições de ensino necessitam extrair informação e conhecimento, presentes em suas bases de dados, para que possam entender o comportamento dos seus alunos. No entanto, muitas dessas instituições armazenam seus dados em formato impresso, implicando na utilização de métodos manuais de tabulação, tornando-se uma ação dispendiosa e inviável quando aplicada em uma base volumosa.

De acordo com a Confederação Nacional da Indústria (2013), a evasão escolar de cursos técnicos de nível médio é um problema que afeta as instituições de ensino e as indústrias brasileiras, pois a falta de trabalhadores qualificados compromete 69% das empresas brasileiras das áreas de transformação e extrativa.

Utilizar os dados registrados sobre os alunos e seu comportamento dentro da escola é uma alternativa para obtermos informações individuais e coletivas no que tange ao sucesso ou insucesso escolar. Este raciocínio é uma possibilidade que podemos utilizar tanto no ensino presencial como a distância, de modo que o segundo tipo fica à frente do primeiro por fazer uso de sistemas que registram o comportamento de quem o utiliza.

Os *softwares* utilizados em Ambientes Virtuais de Aprendizagem (AVA) abstraem funcionalidades que muitos dos seus usuários desconhecem os registrados que são armazenados de forma automática. Além de possuírem a capacidade de armazenar dados pessoais solicitados no ato da matrícula, eles também armazenam todo o caminho percorrido pelo aluno dentro da plataforma, desde a quantidade de acessos, realização de atividades propostas e participações de *chats* e fóruns.

Assim, a partir desses dados registrados nos AVA, surge a importância de conhecermos e entendermos como as técnicas de MD podem ser utilizadas como estratégia de caráter preventivo contra a evasão escolar de cursos presenciais ou a distância.

Diante disso, buscamos responder a seguinte pergunta: como identificar precocemente grupos de alunos propensos a evadir-se de cursos técnicos a distância, partindo da construção e análise de modelo(s) preditivo(s) obtido(s) das etapas e técnicas de MD?

Justificamos nossa pesquisa por, até o presente momento não existir nenhum estudo realizado sobre a evasão escolar dos Cursos Técnicos de Nível Médio Subsequente, na modalidade EAD do Instituto Federal de Educação Ciência e Tecnologia do Piauí (IFPI), no qual a Coordenação Geral tenha utilizado a metodologia

CRISP-DM sobre os dados escolares dos seus alunos por perceber que as técnicas de MDE podem contribuir para a forma de pensar e elaborar estratégias no combate a evasão escolar.

1.2 OBJETIVOS DA PESQUISA

No âmbito do Instituto Federal do Piauí evidencia-se a necessidade de reafirmar ações preventivas de controle e redução da evasão, pois de acordo com Silva (2016), em determinados cursos a distância esse fenômeno compromete mais de 50% das turmas.

Com isso é necessário buscar estratégias ou metodologias que permitam maior fidedignidade/exatidão partindo da importância de saber distinguir entre dados, informação e conhecimento. Entender a diferenciação entre esses elementos pode contribuir para o entendimento sob a ótica da literatura, quando queremos agregar valor ao processo de tomada de decisão sobre um determinado tema, em especial na busca pelo controle da evasão.

Diante do exposto, o presente trabalho tem como objetivo geral adaptar um processo de MD para estimar o risco de evasão de alunos de Cursos Técnicos de Nível Médio a Distância do IFPI utilizando as etapas propostas pela metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*).

Para alcançarmos o objetivo geral, elencamos os seguintes objetivos específicos:

- a) Entender a descoberta de conhecimento sobre dados de grande volume a partir da metodologia CRISP-DM como paradigma de mineração de dados educacionais;
- b) Aplicar a metodologia CRISP-DM para identificar as características comuns de alunos que evadem dos Cursos Técnicos de Nível Médio a Distância do IFPI, para intervenção preventiva da coordenação gestora;
- c) Avaliar o risco de evasão nos Cursos Técnicos de Nível Médio na modalidade EAD do IFPI, partindo de dados acadêmicos de ingressantes no ano de 2013

1.3 ESTRUTURA DA DISSERTAÇÃO

Este trabalho está organizado em cinco capítulos correlacionados. No primeiro capítulo apresentamos a contextualização e o problema da pesquisa, a justificativa, o objetivo geral e os objetivos específicos, fundamentais inicialmente para a compreensão das discussões que se seguem.

O segundo capítulo apresenta o referencial teórico norteador do tema, embasado em publicações que abordam sobre: o fenômeno evasão, entendimento geral sobre a descoberta de conhecimento em base de dados, MDE e a metodologia CRISP-DM. A matriz teórica sustenta-se em Fayyad, Piatetsky e Smyth (1996), Chapman (2000), Dore e Lüsche (2011), Marques (2014), Censo EAD (2015) e Veloso (2015).

Em seguida, o terceiro capítulo expõe a caracterização da pesquisa e a aplicação das fases da metodologia CRISP-DM, descrevendo as tarefas realizadas em cada fase sobre a amostra pesquisada.

O quarto capítulo aborda as discussões dos resultados da avaliação após a etapa da modelagem. As considerações finais desta dissertação serão expostas no quinto capítulo, cuja pretensão é despertar e fazer um chamamento no quanto a MDE pode contribuir para uma análise reflexiva de dados educacionais.

2 REFERENCIAL TEÓRICO

Por servir de base para uma pesquisa, a fundamentação teórica é o momento em que o autor constrói argumentações e demonstrações para mostrar que o problema e as variáveis analisadas estão correlacionados. Em termos teóricos e empíricos, ela também serve para auxiliar na escolha dos instrumentos necessários para a etapa da coleta dos dados.

2.1 EVASÃO ESCOLAR: PERCEPÇÕES E DIMENSÕES

A evasão tem se constituído, ao longo das últimas décadas, como temática desafiadora no sentido de alcançarmos pontos ‘consensuais’ que possam orbitar em torno de seus aspectos tanto conceituais quanto de ordem prática. Todavia, as percepções que foram construídas assinalam que a evasão é um fenômeno que causa impacto negativo na Educação e no desenvolvimento de um país, considerando que além de minimizar a oportunidade de estudo de outros possíveis alunos, maximiza o percentual de profissionais com pouca ou nenhuma qualificação no mercado ou mundo produtivo.

Nessa perspectiva, Veloso (2015) aponta a evasão escolar como sendo um problema educacional que atinge instituições de ensino da rede pública e privada, causando prejuízo para a sociedade, destacando, sobretudo tratar-se de um fenômeno de difícil definição por ser um resultado da interação ou combinação de vários fatores como a proposta de ensino, a avaliação, a interação entre os atores do cenário escolar, o contexto social e familiar do aluno.

Para Figueiredo e Salles (2017), a literatura apresenta um quadro de conceitos diversificados que gera ambiguidade para quem faz análise temática da evasão escolar. Essa ambiguidade está associada a diversas situações como a retenção e repetência do aluno na escola, a saída do aluno da instituição ou do sistema de ensino, a não conclusão de um determinado nível de ensino, o abandono da escola e posterior retorno ou que nunca ingressaram em um determinado nível de ensino (DORE; LÜSCHER, 2011).

Percebemos em nossas leituras que ainda não há uma concordância universal entre os teóricos educacionais quando o assunto é definir o que é a evasão escolar.

Porém, de acordo com Rodriguez (2011), o Ministério da Educação (MEC) define a evasão como sendo a saída definitiva do curso de origem sem conclusão. Nesse aspecto, ousamos abrir um parêntese para uma proposição reflexiva¹, que talvez justifique a imprecisão, lacunas e ambiguidade ao se tratar sobre o termo Evasão: onde estaria a diferença entre desistir de um curso e evadir? Não seria a evasão do Sistema e não do curso? Não deveríamos tratar de desistência do curso e não a evasão deste?

No entanto, entendemos que a evasão não deve ser tratada como um “saldo de caixa”, ou seja, baseado na diferença entre o número de alunos matriculados e o número de concludentes, mas na identificação precoce de elementos que apontam características de discentes que não concluirão um determinado curso.

De fato, ao se analisar o comportamento de um aluno com base em suas características enquanto indivíduo social, existe um grande fator de peso a ser considerado, o que inclui destacar a forma como as instituições de ensino estrutura, organiza e desenvolve seu banco de dados de forma a prever precocemente a identificação dos possíveis estudantes com perfil para a evasão, o que por sua vez reverbera sobre sua gestão estratégica e de planejamento. Nesses termos, convém destacar que o planejamento é a ferramenta para administrar as relações com o futuro, o que representa uma aplicação específica do processo de tomar decisões.

Além de corroborar com os autores supracitados, Silva Filho (2007), entende que a evasão não é caracterizada por um único fator, ao tempo em que enfatiza que as instituições de ensino mencionam como principal razão a falta de condições financeiras por parte do estudante, para continuar os estudos.

Nessa mesma linha de pensamento, Moraes (2010) reforça que para suprir a necessidade financeira, os estudantes enfrentam longas jornadas de trabalho em oposição ao cansaço, acarretando na preferência pelos recursos financeiros ao invés da sua qualificação para alcançar novas oportunidades profissionais.

Além disso, o autor atribui à evasão, outros fatores como a dificuldade de aprendizagem e o engano na escolha do curso. Para ele, o primeiro fator provém dos alunos estarem habituados com processos de ensino limitados a técnicas de memorização, confrontando com a necessidade de se desenvolver um cidadão capaz

¹ A temática suscita tal reflexão, contudo, não nos aprofundaremos, por não se tratar do foco desta pesquisa, mas merecedora de outras reflexões posteriores.

de construir sua própria opinião. Enquanto o segundo, parte do equívoco na escolha da profissão por parte do aluno.

De uma forma mais específica, outros pesquisadores entendem e concordam que a evasão escolar pode ser analisada com base em dimensões econômicas, sociais, educacionais, de tempo, de localização e tecnológica. O **Quadro 1** mostra alguns destes.

Quadro 1 - Causas da evasão escolar por dimensão

DIMENSÕES	VARIAVEIS	REFERÊNCIAS
Econômica	Problemas financeiros	GAIOSO (2005)
	Necessidade de trabalhar	DORE e LÜSCHER (2011)
Social	Idade, sexo, raça e contexto social	CARVALHO (2003); GAIOSO (2005); MARQUES (2011)
	Grau de instrução familiar	GAIOSO (2005)
Educacional	Deficiência na educação básica, especialmente em Matemática e Língua Portuguesa	GAIOSO (2005); MARQUES (2011)
	Repetência escolar	DORE e LÜSCHER (2011)
Tempo	Disponibilidade de tempo	DORE e LÜSCHER (2011); GAIOSO (2005)
Localização	Instabilidade residencial e mudança de cidade	DORE e LÜSCHER (2011)
	Incompatibilidade entre local da escola e o local da moradia, falta de transporte até a escola	DORE e LÜSCHER (2011); GAIOSO (2005)
Tecnológica	Pouco domínio da tecnologia, tecnologia substituindo o professor, dificuldade de acesso às tecnologias, dificuldade de conexão	ZORDAN (2012)

Fonte: Dados da pesquisa (2017).

Assim como nos cursos presenciais, a evasão em cursos a distância é um problema que preocupa as instituições de ensino, pois a pouca produção científica sobre a temática provoca a necessidade de ampliarmos as discussões sobre o tema, tendo em vista que os estudos e práticas para redução da evasão não são suficientes (ZORDAN, 2012).

Estatisticamente, de acordo com o Censo EAD de 2015, a evasão nos cursos a distância é maior do que nos cursos presenciais. Além disso, numa variação de 26% a 50%, em cursos regulamentados totalmente a distância essa taxa equivale a 40%.

De acordo com referido Censo, as Instituições apontam a dimensão tempo como o mais influente no fenômeno da evasão, seguido do fator finanças, falta de adaptação à modalidade ou à metodologia do curso, e escolha errada do curso.

Dos fatores que foram identificados para justificar a evasão de cursos a distância, a falta de tempo é a argumentação menos aceitável. No entanto, acreditamos que este argumento esteja associado à dificuldade de organização e disciplina na administração do horário destinado para os estudos. Afinal de contas, disciplina e tempo são características fundamentais e inerentes ao perfil do aluno que deseja uma formação a distância.

Embora o Censo EAD de 2015 aborde a evasão comparando ensino a distância com o presencial, não identificamos um detalhamento no que remete ao quantitativo por nível de ensino existente em cada modalidade supracitada. Contudo, acreditamos que a análise sobre a evasão escolar por nível de ensino pode nos dar um panorama mais preciso de fatores comuns e distintos que levam um aluno a não concluir um curso dentre os diversos níveis de ensino em caráter presencial ou a distância.

Nessa mesma percepção, Dore e Lüscher (2011) consideram importante sabermos o nível escolar em que ocorre a evasão. Para elas, a evasão é significativamente diferente se ela sucede no nível fundamental, médio, na educação de adultos ou na educação superior. As autoras reforçam exemplificando que a evasão de estudantes que abandonam o ensino superior pode ser inserida na categoria dos que não continuaram a educação que iniciaram, mas que em termos de escolaridade, encontram-se em situação muito diferente daqueles que não concluíram ao menos a educação obrigatória.

Além disso, Dora e Lüscher (2011), consideram um desafio pesquisar sobre a evasão escolar no ensino técnico no Brasil, argumentando quanto à falta de informações sobre o assunto no que tange a referenciais teóricos e empíricos. A fragilidade nestes aspectos contribui para o quantitativo de dificuldades encontradas na construção de indicadores coerentes à investigação desse fenômeno.

Para Lobo (2012), a evasão pode ser comparada às doenças silenciosas e assintomáticas sem indicadores do que irá acontecer ou que esteja acontecendo. De fato, a evasão é uma ameaça que compromete bem mais que a capacidade das instituições de ensino no que se refere a atender às exigências do mercado de trabalho; ela força a necessidade de buscarmos constantemente ações preventivas contra este fenômeno.

Quanto a estas ações Dore e Lüscher (2011), expressam que a complexidade no andamento da evasão escolar também requer soluções complexas envolvendo diversos agentes sociais. Além disso, as autoras afirmam que a maior parte das pesquisas propõe a prevenção da evasão escolar, sob o aspecto da identificação precoce e o acompanhamento individual daqueles que estão em situação de risco.

De acordo com Veloso (2015), a identificação de alunos propensos a se evadirem pode ser caracterizada a partir de vários modelos matemáticos, estatísticos ou de aprendizagem de máquina. Logo, a descoberta do conhecimento sobre banco de dados e a aplicação de uma metodologia de MD pode ajudar em situações de matrículas em andamento no que concerne a estimar e identificar alunos matriculados em cursos presenciais ou a distância, com maior risco de evasão.

Diante do exposto, percebemos que é relevante o nosso estudo, por entendermos que estimar a probabilidade de um aluno evadir-se de um curso baseado em modelo(s) não linear(es) de características explicativas tomadas como motivadoras partindo de dados registrados nos formulários de matrícula requer conhecimento e uso da tecnologia sobre dados volumosos.

Com isso, usar o conhecimento sobre banco de dados como estratégia para analisar e conhecer o estudante na sua individualidade e na coletividade com outros estudantes matriculados independe das definições para evasão. Partindo disso, consideramos como evadido, em nosso trabalho, o aluno que nunca teve contato com o curso após a matrícula ou que abandonou o curso em caráter definitivo. Nesse sentido, conceitualmente destacamos evasão como a falta de contato do aluno com o curso após efetivar a sua matrícula.

Desta forma, nosso trabalho vislumbra as potencialidades que um banco de dados educacionais pode revelar sobre os discentes considerando que o conhecimento sobre esses dados pode ser um ponto de partida no que tange a buscar estratégias contra o fenômeno evasão.

De acordo com Date (2003), os bancos de dados são sistemas computadorizados preparados para armazenar registros e permitir que seus usuários acessem e atualizem os dados ali presentes, quando necessário, mediante a sua importância. Com isso, podemos dizer que ao tempo em que as Instituições de Ensino coletam e armazenam dados cadastrais e comportamentais de seus alunos, limitando-se apenas em analisar o perfil dos discentes sob resultado do cálculo da frequência relativa para cada variável analisada, as mesmas devem perceber e conhecer

ferramentas e metodologias que podem otimizar e apontar com maior precisão a estimativa ensejada.

Portanto, a adesão de mecanismos automatizados e metodologias que viabilizam a detecção precoce de um aluno ou grupo de alunos com risco de evasão, também é uma condição importante para reduzir o problema.

2.2 MINERAÇÃO DE DADOS

A partir do momento em que a informática tomou conta de nossas vidas, grandes volumes de dados têm sido sistematicamente coletados e armazenados. Para Cardoso e Machado (2008), o surgimento de enormes bases de dados é atribuído ao barateamento de computadores de alto desempenho, e que devido ao seu volume, ultrapassa a capacidade humana na extração do conhecimento.

Nesse contexto, percebemos a necessidade da utilização de técnicas e ferramentas tecnológicas que permitam a organização, análise e a extração de conhecimento sobre dados volumosos, de maneira total ou parcialmente automatizadas.

Com isso, no final da década de 80, foi proposta a MD ou *Data Mining* (DM), por profissionais de organizações que começaram a se preocupar com os grandes volumes de dados armazenados, subutilizados ou nunca utilizados pelos seus detentores.

Considerada atualmente uma das tecnologias mais promissoras quando o assunto é buscar conhecimento em dados volumosos, a MD é o termo utilizado para entendermos a realização de etapas e tarefas com o objetivo de descobrir o conhecimento sobre modelos obtidos dentro de um contexto que requer tomada de decisão diante de um problema.

Por haver várias definições utilizadas na literatura da Ciência da Informação, é importante entendermos os conceitos relacionados aos termos dado, informação e conhecimento.

Para Miranda (1999), o dado é um conjunto de registros qualitativos ou quantitativos que se transformam em informação, após ser organizado, agrupado, categorizado e padronizado adequadamente. De uma forma mais ampla, Davenport (1998, p. 18 *apud* VALENTIM, 2002, p. 2) diz que um dado é a representação das “observações sobre o estado do mundo”.

No que remete ao entendimento sobre o que é informação, Barreto (1996, p. 2) diz que “são estruturas com a competência de gerar conhecimento no indivíduo, em seu grupo, ou na sociedade”. De uma forma mais harmônica com este trabalho, Miranda (1999) define informação como dados organizados de modo a gerar impacto na tomada de decisão.

Quanto à compreensão do termo conhecimento, Barreto (1996, p. 2) refere-se como “alteração no estado cognitivo do indivíduo, isto é, no seu estoque mental de saber acumulado, proveniente de uma interação positiva com uma estrutura de informação”. De uma forma mais sofisticada, Davenport expõe o seu entendimento sobre conhecimento como a mistura de experiências, valores e informações contextuais que proporcionam uma estrutura para a avaliação e a incorporação de novas experiência e informações (1998, p. 18 *apud* VALENTIM, 2002).

Portanto, diante do exposto, podemos dizer que o conhecimento é a mudança intelectual de um ou mais indivíduos após tomar decisão diante de dados sistematizados.

Por outro lado, a descoberta de padrões ou modelos depende da escolha e da disponibilidade de acesso aos dados que podem respaldar as respostas que se deseja alcançar sobre um determinado problema. Para isso, os dados deverão ser selecionados, integrados, limpos e padronizados para que sejam submetidos ao processo de mineração.

De acordo com Camilo e Silva (2009), a MD é classificada de acordo com as tarefas a serem realizadas para se alcançar os objetivos previamente estabelecidos. Dentre as diversas tarefas a serem realizadas para a extração do conhecimento utilizando a MD, as mais comuns são: a descrição, a classificação, a estimação ou regressão, a predição, o agrupamento e a associação, complementam os autores.

A primeira tarefa é utilizada para descrever padrões revelados pelos dados, oferecendo uma possível interpretação para os resultados obtidos e comprovar a influência de variáveis.

A segunda identifica a classe a que um determinado registro pertence. É a etapa em que o modelo “aprende” como classificar um novo registro para categorizá-lo em uma classe.

A regressão, embora seja similar à classificação, é usada quando o registro tem valor numérico e não categórico; assim, podemos estimar o valor de uma variável alvo, analisando os valores das demais.

Já a predição é utilizada como mecanismo para descobrir o valor futuro de um determinado atributo.

Por fim, o agrupamento identifica grupo de dados similares e a associação consiste em identificar quais atributos estão relacionados entre si apresentando-se em forma de implicação lógica.

De acordo com Larose (2005), embora a definição de MD possa transparecer um processo de extração de conhecimento que acontece de forma exclusivamente automática, sabe-se hoje que de fato isso não é verdade, pois os resultados ainda precisam de uma análise humana. Porém, ainda assim, a mineração contribui de forma significativa no processo da descoberta de conhecimento, permitindo aos especialistas concentrarem esforços apenas em partes mais significativas dos dados.

De forma clássica, a MD é a etapa mais importante de um processo mais amplo conhecido como Descoberta de Conhecimento em Bases de Dados (DCBD) ou *Knowledge Discovery in Databases* (KDD).

Fayyad, Piatetsky e Smyth (1996), referem-se ao termo KDD como o processo não trivial da extração de conhecimento dos dados como um todo, enquanto a MD corresponde apenas a uma etapa particular do KDD, na qual a extração de padrões dos dados é realizada através do uso de algoritmos específicos.

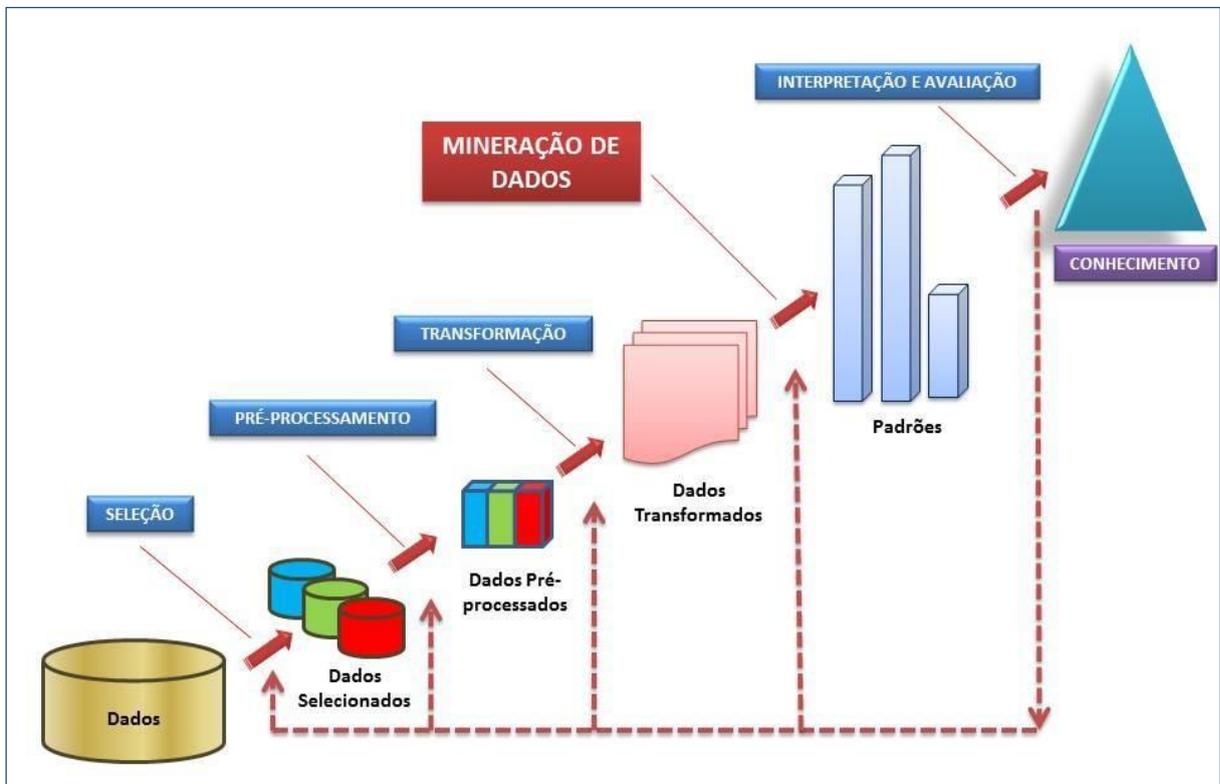
Diante disso, percebemos que o processo de extração de conhecimento sobre dados requer a criação de padrões e modelos no que tange à transformação de dados em informações que comprovam um pressuposto.

Segundo Camilo e Silva (2009), ainda não há um consenso quanto à definição de KDD e *Data Mining*, por serem considerados sinônimos no entendimento de alguns pesquisadores, porém todos concordam que o processo de mineração deve ser iterativo, interativo e dividido em fases.

Para Fayyad, Piatetsky e Smyth (1996), a Descoberta de Conhecimento em Bases de Dados divide-se em etapas que envolvem a preparação dos dados, busca por padrões, avaliação e refinamento, além de ser um processo interativo, por possuir etapas que contêm tarefas e decisões a serem realizadas por um analista humano conhecedor do negócio envolvido, e é iterativo, pois todas as etapas são sistematicamente interligadas.

A **Figura 1** representa o processo da Descoberta do Conhecimento em Bases de Dados, segundo Fayyad, Piatetsky e Smyth (1996). Em seguida, explicitamos o que significa cada uma das etapas.

Figura 1 - Descoberta de Conhecimento em Bases de Dados



Fonte: Fayyad, Piatetsky e Smyth (1996).

- a) Seleção: considerada como a primeira etapa da DCBD, nesta instância é criado um conjunto ou subconjunto de dados que será o foco da descoberta de novos conhecimentos. Ele deve conter as informações necessárias para que os algoritmos de mineração possam alcançar o objetivo do pesquisador.
- b) Pré-processamento: momento em que os dados passam por uma limpeza ou eliminação de ruídos, e que inclui operações básicas para remoção de inconsistências.
- c) Transformação: etapa da formatação necessária para agregar valor semântico às informações ou características úteis para representar os dados da base.
- d) Mineração de dados: aplicação das técnicas de MD usando algoritmos para alcançar os objetos definidos na etapa da Seleção.

- e) **Interpretação e Avaliação:** compreensão dos padrões obtidos incluindo a visualização dos modelos que resumem a estrutura e as informações presentes nos dados juntamente com as medidas técnicas que avaliam.

A utilização de recursos computacionais para a busca de padrões potencialmente úteis e interpretáveis, a partir das grandes bases de dados, pode requerer a repetição parcial ou integral das etapas e tarefas supracitadas, na quantidade de vezes que se fizerem necessárias para se chegar ao objetivo desejado. Para tanto, é necessária a utilização de uma metodologia capaz de cumprir os requisitos operacionais, tornando o processo de Descoberta de Conhecimento em Bases de Dados confiável e gerenciável.

O processo de descobrir conhecimento a partir de dados brutos pode ser estruturado em formato piramidal, onde se começa com uma base elementar e volumosa, finalizando em um ponto mais concentrado e mais valioso. Para Navega (2002), a **Figura 2** representa essa compreensão.

Figura 2 - Pirâmide da Informação



Fonte: Navega (2002).

Nos últimos anos, o uso de Descoberta de Conhecimento em Bases de Dados e técnicas de MD, vem despertando o interesse de pesquisadores no que toca a sua aplicação sobre dados educacionais, denominando-a Mineração de Dados Educacionais (MDE).

Para Costa *et al.* (2012) a MDE é a área emergente que procura desenvolver ou adaptar métodos para serem aplicados em dados de contextos educacionais com o intuito de analisar o comportamento de estudantes e a forma como eles aprendem.

Com isso podemos dizer que a aplicação de técnicas de MD pode ser útil para os gestores educacionais na busca de um acompanhamento mais diferenciado e proativo, principalmente na identificação de alunos ou grupo de alunos propensos à evasão.

2.3 MINERAÇÃO DE DADOS EDUCACIONAIS

A Mineração de Dados Educacionais (MDE) ou *Educational Data Mining* (EDM) é uma área recente de pesquisa. Tem como objetivo, desenvolver métodos para explorar dados de grandes volumes provenientes de uma variedade de fontes relacionadas a um contexto educacional, úteis no processo de tomada de decisão no âmbito escolar (BAKER; ISOTANI; CARVALHO, 2011).

Embora a MDE no Brasil esteja ganhando espaço, internacionalmente ela está bem consolidada desde o primeiro *Workshop, Educational Data Mining*, durante o *20th National Conference on Artificial Intelligence*, em Pittsburg-EUA, no ano de 2005 (MARQUES, 2014).

Conforme Silva *et al.* (2015), a MDE pode ser entendida como o processo de descoberta de conhecimento sobre dados brutos de sistemas educacionais útil a desenvolvedores de *softwares* educacionais, professores, pesquisadores educacionais.

Nessa perspectiva, entendemos que a MDE pode servir de base na construção de sistema de alertas que podem ajudar professores e gestores na identificação de comportamentos em situações que podem levar à reprovação ou à evasão escolar.

Apesar da MDE proporcionar a descoberta do conhecimento útil a educadores, o processo de extração ainda está limitado a profissionais da informática, estatísticos e etc, devido à complexidade de como o processo de extração é realizado.

Quanto aos métodos utilizados na aplicação de técnicas de MD em ambientes virtuais, Garcia *et al.* (2011), ressaltam que difere muito da aplicação das técnicas em outras áreas, pois há casos em que o volume de dados disponíveis é considerado pequeno ao ser comparado com a quantidade de atributos.

Diante do posicionamento dos recém citados autores, a busca por conhecimento em dados volumosos é válida, mesmo que haja limitações quanto às fontes de coletas e à quantidade de componentes da amostra e de atributos. No

entanto, quanto mais informações tivermos à disposição para análise, mais ricos serão os resultados da mineração.

Paralela à situação supracitada, a identificação do quantitativo de publicações relacionados a MDE, e sua aplicação, motivou Machado *et al.* (2015) a fazerem o levantamento desse quantitativo relacionando a MD e a evasão escolar. A pesquisa visava apresentar um estudo bibliométrico de artigos publicados que abordavam a utilização das técnicas de MD no contexto da educação e aplicada ao problema da evasão escolar.

Enquanto Machado e seus colegas (2015) têm a concepção de que o quantitativo de publicações sobre a MDE é uma forma de expressar a necessidade de se conhecer a aplicabilidade da MD envolvendo a evasão escolar, em 2014, Adeodato, Santos Filho e Rodrigues publicaram um artigo utilizando a regressão logística na MD para avaliar a qualidade da educação secundária privada no Brasil a partir das bases do Exame Nacional do Ensino Médio (ENEM) e do Censo Escolar disponibilizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP).

De acordo com os autores, o Exame Nacional do Ensino Médio captura informações socioeconômico e culturais do perfil dos participantes, e partindo desse entendimento, os mesmos utilizaram a metodologia CRISP-DM para caracterizar o que consideram como “boa” escola. Para isso, os mesmos fizeram uso da Regressão Logística para produzir um classificador de pontuação. A **Equação 1** permitiu a determinação do *score* de propensão de sucesso das escolas analisadas.

Equação 1 - Regressão logística

$$\log \left\{ \frac{\pi(x)}{1 - \pi(x)} \right\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Fonte: Dados da pesquisa (2017).

Ressaltamos que a variável π é a probabilidade de um evento e β corresponde ao valor que influencia o atributo.

De acordo com a ideia dos autores supracitados, podemos reescrever a Equação 1 e obtermos a **Equação 2**.

Equação 2- Função logística

$$\pi(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}$$

Fonte: Dados da pesquisa (2017).

A função logística é muito útil para quem trabalha com cálculos estatísticos. De fato, através dela é possível mensurar a probabilidade da ocorrência de um evento por sua imagem variar de 0 a 1. Com isso, entendemos que a função logística pode ser aplicada em problemas educacionais como estimar o risco de evasão escolar, considerando x_i , (com $i = 0, 1, 2, 3, 4, 5, \dots, p$) como variáveis que podem justificar a propensão de um aluno a evadir-se de uma instituição.

Em outro trabalho, Barbosa *et al.* (2014) enfatizaram a aplicabilidade da descoberta de conhecimentos em banco de dados educacionais utilizando a metodologia CRISP-DM na identificação das causas da evasão na EAD de cursos técnicos, graduação e de pós-graduação. Segundo os autores, a metodologia dá um panorama compreensivo do processo de MD, semelhante a outras estruturas já existentes, mas que se destaca por ser considerada como completa, organizada e com estrutura de fácil entendimento ou revisão de todo o projeto.

Diante disso, nosso trabalho visa utilizar as técnicas de MDE seguindo as etapas da metodologia CRISP-DM para estimarmos o risco de evasão de alunos de Cursos Técnicos de Nível Médio a Distância do IFPI. Este entendimento parte do pressuposto de que os dados informados no questionário socioeconômico podem contribuir na obtenção da probabilidade de um aluno se evadir antes do primeiro contato com o curso. Dessa forma, estratégias ou ações pedagógicas por parte da coordenação poderão ser tomadas de forma preventiva.

2.4 CRISP-DM

A metodologia CRISP-DM ou *Cross-Industry Standard Process of Data Mining* (Processo Padrão Inter-Indústrias para Mineração de Dados) foi criada durante a década de 90 e sua origem se deve principalmente à necessidade da elaboração de modelos com foco na qualidade através da padronização de conceitos e técnicas,

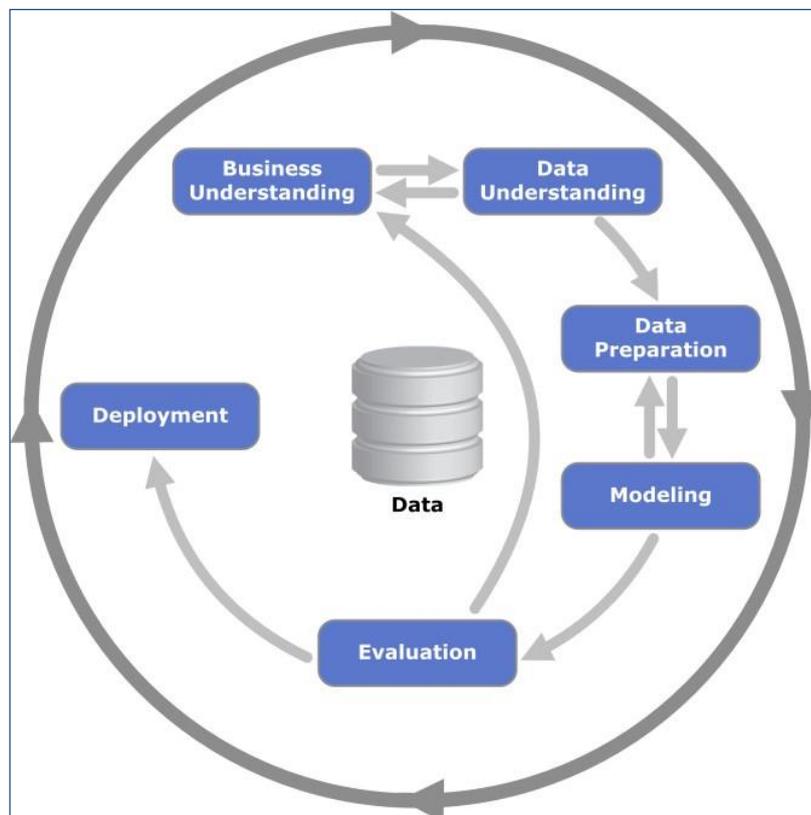
busca de informações e tomada de decisões. Ela propõe um modelo de MD de forma abrangente.

Sua criação propõe auxiliar os responsáveis pelos processos, desde o planejamento até a execução da MD, passando pela especificação do processo até a apresentação dos resultados alcançados.

Segundo Chapman (2000), a metodologia CRISP-DM é composta por 6 fases, organizadas de maneira cíclica, cujo fluxo é não unidirecional, possibilitando ir e voltar entre as suas fases e tarefas: *Business Understanding* (Entendimento do Negócio), *Data Understanding* (Entendimento dos Dados), *Data Preparation* (Preparação dos Dados), *Modeling* (Modelagem), *Evaluation* (Avaliação) e *Deployment* (Implementação do Modelo).

A **Figura 3** representa a sequência de cada uma das fases que compõem a metodologia.

Figura 3 – Metodologia CRISP-DM



Fonte: Chapman (2000).

De acordo a **Figura 3**, percebemos que a metodologia CRISP-DM define um processo de MD em formato cíclico e que, implicitamente, cada fase é composta de

tarefas que deverão ser realizadas para que seja determinada qual fase ou tarefa será a próxima a suceder.

Vale ressaltar que embora a referida metodologia tenha um formato padrão de fases do processo de MD, a mesma não impede a possibilidade da inserção de novas amostras, variáveis ou atributos a serem minerados.

A seguir, definimos sucintamente cada uma das fases da metodologia CRISP-DM, de acordo com Chapman (2000):

- a) Entendimento do Negócio: fase de delimitação dos objetivos do projeto e premissas necessárias para a perspectiva do negócio. Nesta etapa o conhecimento é convertido em um problema de MD e na busca por uma solução. É elaborado um plano estratégico preliminar que contribua para o alcance dos objetivos definidos previamente;
- b) Entendimento dos Dados: momento de familiarização com os dados e identificação de problemas na qualidade dos mesmos. É nesta fase que são construídas as primeiras percepções e realizadas formulações de hipóteses sobre o que os dados podem revelar;
- c) Preparação dos Dados: fase que abrange todas as atividades para a construção final de dados para a etapa da modelagem. Neste estágio, ocorre a seleção e integração de tabelas, limpeza dos dados brutos iniciais, construção de gráficos e elaboração do dicionário de variáveis onde são classificadas em numéricas ou categóricas.
- d) Modelagem: nesta fase várias técnicas de modelagem são selecionadas e aplicadas e os parâmetros destas técnicas são ajustados para sua otimização. Tipicamente, são técnicas distintas para os mesmos tipos de problemas de MD, sendo que algumas têm necessidades específicas na transformação do dado;
- e) Avaliação: a obtenção de um modelo ou vários modelos deve apresentar qualidade para a análise dos dados. Antes da construção final do modelo, é importante avaliar e rever os passos executados para sua construção como

forma de ter certeza de que a sua execução permitirá o alcance dos objetivos. Mas a criação do modelo não é o fim do projeto;

- f) Implantação: fase em que o conhecimento adquirido necessita ser organizado e apresentado a um contexto em que possa ser utilizado.

Com isso, de acordo com as fases supracitadas percebemos que a metodologia CRISP-DM pode ser aplicada em diferentes contextos que lidam com dados volumosos nos levando a crer na sua potencialidade em direcionar a descoberta do conhecimento em dados educacionais.

Assim, na intenção de descobrirmos o risco de evasão sobre os dados da nossa pesquisa, dedicamos o próximo capítulo às fases e tarefas que foram realizadas ao adaptarmos a metodologia CRISP-DM.

3 METODOLOGIA DA PESQUISA

Tendo como ponto de partida uma dúvida ou um problema a ser resolvido, o pesquisador consegue buscar uma saída ao fazer uso de um ou mais métodos científicos. A metodologia da pesquisa tem por objetivo apresentar uma sequência lógica de etapas a serem seguidas para alcançar os objetivos propostos, com resultados coerentes (WAZLAWICK, 2008).

Com isso, podemos dizer que ao seguirmos as etapas da metodologia de uma pesquisa, estaremos respeitando as normas necessárias para que o conhecimento produzido não perca sua validade.

3.1 CARACTERIZAÇÃO E CONTEXTUALIZAÇÃO DA PESQUISA

A pesquisa científica é fruto de uma investigação minuciosa, visando resolver um problema e para isto é indispensável selecionar o método a utilizar (FONSECA, 2002).

Para Silva (2005), existem várias formas de classificarmos uma pesquisa, e que a mesma pode ser enquadrada em várias classificações simultaneamente, desde que atenda às características inerentes de cada tipo.

Diante disto, a nossa pesquisa é considerada como documental e aplicada. De fato, para Gil (1996) a pesquisa documental é elaborada a partir de materiais que não receberam tratamentos analíticos ou que podem ser reformulados em consonância com os objetivos da pesquisa a se realizar.

Oliveira (1998) complementa a ideia de Gil (1996) afirmando que a pesquisa documental possibilita o encontro de uma série de informações para certificar a presença ou não de uma determinada hipótese.

Além disso, Appolinário (2011, p. 146) diz que a pesquisa aplicada é executada com a intenção de “resolver problemas ou necessidades concretas e imediatas”. Com isso, podemos dizer que a pesquisa aplicada pode gerar conhecimento devido à aplicação prática para tentar solucionar um problema específico.

Quanto à abordagem do problema, a pesquisa é do tipo quantitativa, por necessitar da mensuração de hábitos e atitudes em um universo, por meio de uma amostra que o represente estatisticamente (DENZIN; LINCOLN, 2005).

Para Alves-Mazotti (2004), a pesquisa quantitativa é caracterizada como dedutiva, cuja comprovação é feita através de testes de teorias e predição, focada na quantidade, e no isolamento de variáveis com a análise dos dados utilizando a estatística.

Nesse mesmo entendimento, Freixo (2012) considera a pesquisa quantitativa um processo sistemático de coleta de dados observáveis e quantificáveis baseados na observação de fenômenos e acontecimentos que existem independentemente do investigador.

Para Reis (2010), este método considera que todos os dados são quantificáveis e podem ser traduzidos em números, opiniões e informações para serem classificados e analisados utilizando métodos estatísticos.

Além disso, é descritiva por descrever a relação existente entre a combinação das variáveis da amostra e o fenômeno evasão, utilizando MD.

Para alcançarmos os objetivos desta pesquisa, aplicamos a metodologia CRISP-DM em uma base de dados de 198 alunos matriculados no Curso Técnico de Nível Médio Subsequente em Segurança do Trabalho com modalidade EAD, ofertado no Instituto Federal de Educação, Ciência e Tecnologia do Piauí – *Campus* Teresina Zona Sul, que ingressaram em 2013.2 por meio de processo seletivo em Matemática e Português. O referido curso possui matriz curricular composta por 24 disciplinas totalizando 1.200 horas, distribuídas em 4 módulos de 300 horas, onde no 1º módulo o aluno construirá as competências básicas para atuar como Técnico em Segurança do Trabalho em empresas públicas e privadas.

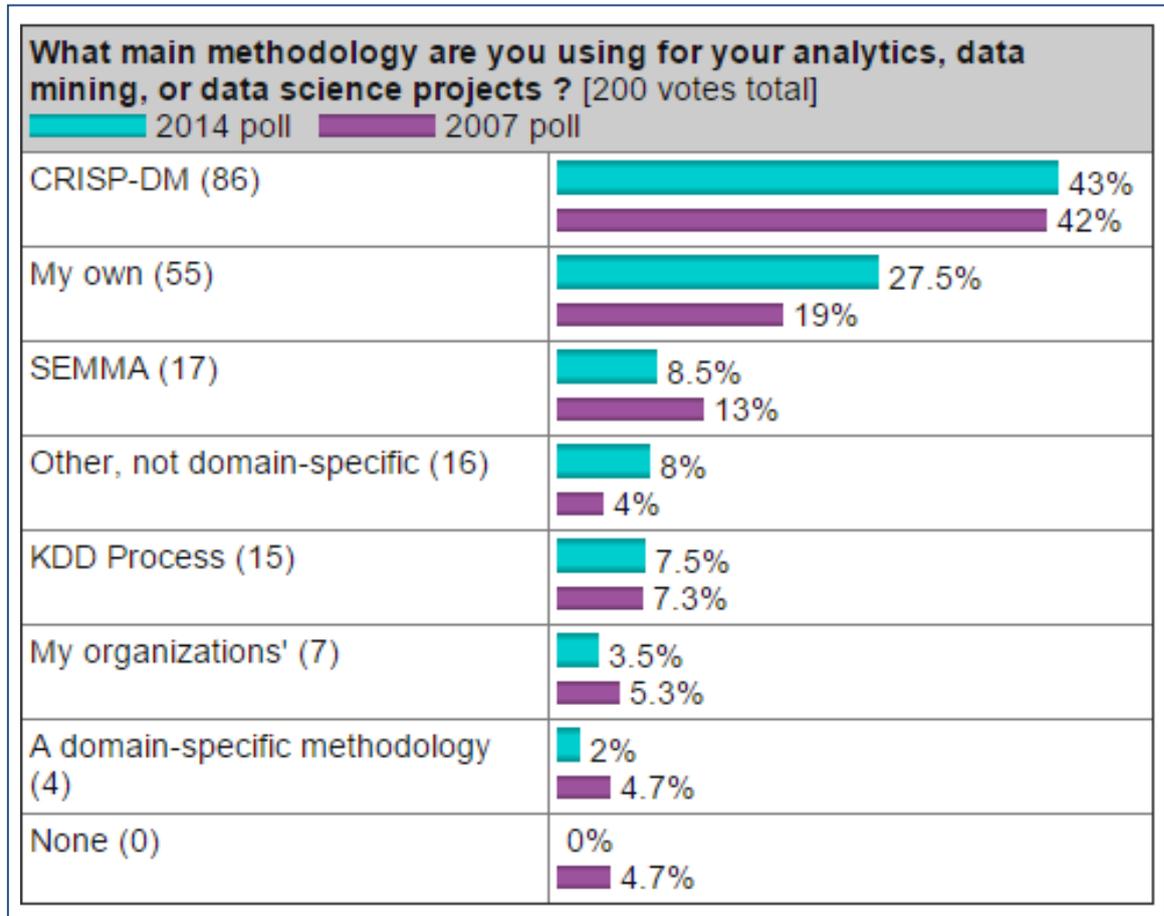
Para compor a base de dados desta pesquisa, tomamos como fonte os condicionantes educacionais sociais e econômicos de cada aluno com a finalidade de obtermos um parâmetro de alerta na identificação dos propensos a se evadirem de cursos técnicos na modalidade a distância do IFPI, utilizando técnicas de MD.

Neste trabalho, escolhemos o CRISP-DM como metodologia por ser considerado atualmente o padrão de maior aceitação entre especialistas, de acordo com o *site*² de KDD chamado *Kdnuggets*. O referido site é líder em *Big Data* e *Data Mining* e coordenado por Gregory Piatetsky-Shapiro, um dos principais pesquisadores na área. Ainda segundo o *site*, a metodologia CRISP-DM é a mais utilizada para alcançar objetivos na resolução de problemas envolvendo *data mining*, sendo a

² <http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>

preferida por 43% dos profissionais que trabalham com MD, de acordo com a **Figura 4**.

Figura 4 - Preferência pela Metodologia CRISP-DM



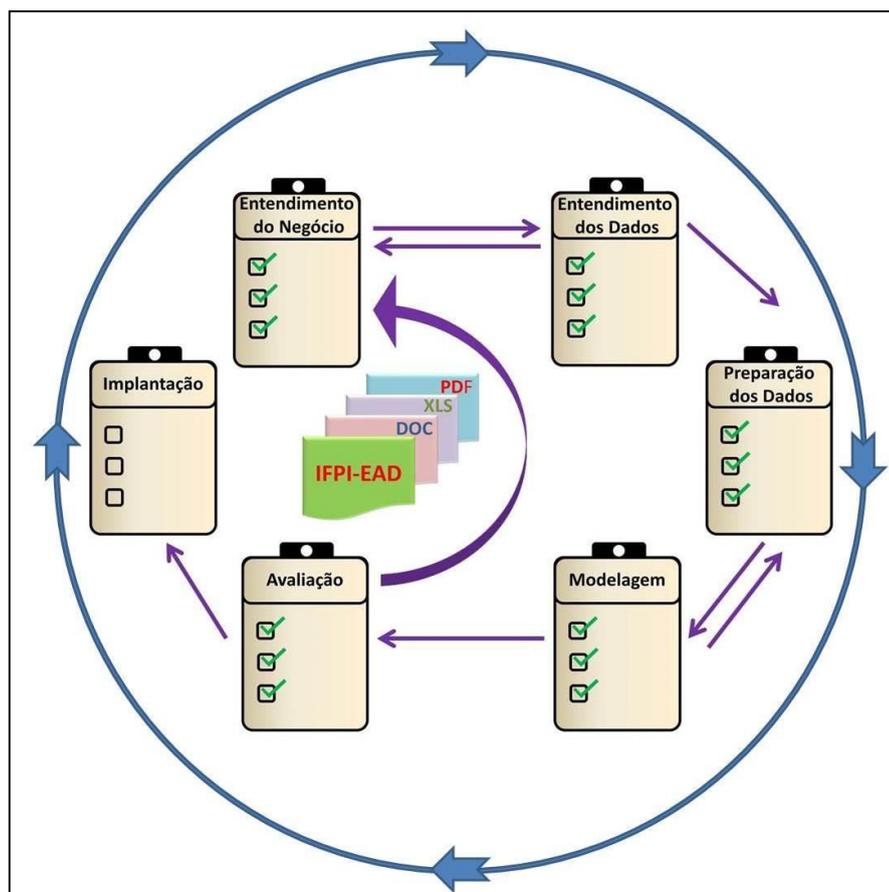
Fonte: *Kdnuggets* (2014).

3.2 APLICAÇÃO DA METODOLOGIA CRISP-DM

Por ter um formato de ampla iteratividade com as suas fases, as quais são flexíveis e sem ordenamento, e por facilitar retornos e avanços com o encerramento de cada fase ou tarefa, a metodologia do presente trabalho foi descrita de maneira a atender as etapas fixas da metodologia CRISP-DM.

Justificamos o uso desta por ser possível detalharmos o caminho percorrido para conseguirmos estimar o risco de evasão escolar em cursos a distância, partindo de uma base de dados com informações cadastrais e comportamentais de alunos do IFPI, conforme a **Figura 5**.

Figura 5 - Metodologia CRISP-DM aplicada a EAD do IFPI



Fonte: Figura adaptada de CHAPMAN (2000).

Vale lembrar que a base de dados utilizada nesta pesquisa é oriunda da compilação das informações registradas nos questionários socioeconômicos dos alunos do curso apontado anteriormente, integradas ao desempenho nas disciplinas cursadas por módulo. Quanto a esses questionários, observamos que os mesmos estavam em formato impresso, utilizados como ficha de matrícula, cujos itens são os mesmos do questionário para cursos presenciais do IFPI.

Ao solicitarmos os dados de registros gravados e quantificados na participação de fóruns, *chats* e etc, quando a plataforma *Moodle* é utilizada pelos discentes, fomos informados que não seria possível, por não serem armazenados por comprometer a capacidade no servidor do IFPI.

3.2.1 Entendimento do Negócio

Fase em que um problema do negócio é analisado como problema de MD aliado à delimitação de objetivos a serem alcançados conforme um plano estratégico.

Baseado nisso e no contexto do local da pesquisa temos a seguinte situação: o IFPI é uma instituição *multicampi* e que oferta cursos técnicos de nível médio e superior nas diversas modalidades de ensino. Dentre estes destacamos os cursos EAD, por terem maior alcance, devido ao formato da modalidade e por contribuírem para a democratização do ensino público.

O IFPI, a cada ano, sempre busca criar e divulgar novas ofertas de vagas de cursos técnicos a distância, cujo processo de preenchimento é feito por meio de processo seletivo, mas o mesmo não consegue manter uma proporção de concluintes equivalente ao número de alunos matriculados inicialmente, evidenciando a fragilidade em lidar com a evasão escolar.

A elevada taxa de evasão não significa apenas a perda dos alunos, mas também de investimento financeiro e humano, necessários para a efetivação de um ensino público de qualidade e que atenda às demandas do mercado de trabalho.

Com isso, considerando a proposta deste trabalho e a justificativa pelo qual escolhemos a metodologia CRISP-DM, foram realizadas as tarefas norteadoras que caracterizam a primeira fase da estrutura do percurso metodológico descritas no **Quadro 2**.

Quadro 2 - Tarefas do Entendimento do Negócio

TAREFA	DESCRIÇÃO
01	Caracterização do negócio: evasão no curso técnico em EAD do IFPI
02	Definição dos objetivos a serem alcançados para estimar o risco da evasão em cursos técnicos na modalidade EAD
03	Plano estratégico: transformar o negócio em problema de MD utilizando estatística descritiva e técnicas de MDE utilizando os dados cadastrais e comportamentais dos alunos da amostra, matriculados no período 2013.2 aprovados em processo seletivo em matemática e português. Consideraremos a evasão como a ação realizada pelo aluno em abandonar um curso logo após a realização da sua matrícula ou contato com o curso por algum intervalo de tempo com obtenção de média aritmética das médias finais das disciplinas ofertadas no primeiro módulo igual a zero

Fonte: Dados da pesquisa (2017).

3.2.2 Entendimento dos Dados

Fase em que os dados são coletados, descritos e explorados, para que seja verificada a sua qualidade e a relação com o problema do negócio. Através desta etapa foi possível delimitarmos quais as fontes de dados seriam utilizadas em nosso

estudo. Das fontes disponíveis e analisadas foram consideradas: o resultado do processo seletivo para ingresso, o questionário socioeconômico de cada aluno matriculado e o mapa de notas de todas as disciplinas ofertadas durante o curso, distribuídas por módulos.

O **Quadro 3**, detalha o caminho percorrido para acesso às informações aludidas acima.

Quadro 3 - Tarefas do Entendimento dos Dados

TAREFA	DESCRIÇÃO
01	Busca no portal do IFPI sobre o resultado final do processo seletivo para cursos técnicos a distância com ingresso no período 2013.2, localizado em formato de mídia com extensão PDF
02	Solicitação de acesso à lista de alunos matriculados após aprovação no processo seletivo, junto ao controle de disciplina do <i>campus</i> .
03	Solicitação de acesso aos dados registrados nos questionários socioeconômicos dos alunos, preenchidos em formulário impresso no ato da matrícula.
04	Solicitação de acesso ao mapa de notas das disciplinas ofertadas/cursadas em cada módulo do referido
05	Formulação de hipótese: a pontuação obtida no processo seletivo, os dados informados nos questionários socioeconômicos e mapa de notas das disciplinas ofertadas em cada módulo podem revelar conhecimento sobre a evasão no Curso Técnico em Segurança do Trabalho

Fonte: Dados da pesquisa (2017).

3.2.3 Preparação dos Dados

A preparação dos dados implica na construção de um conjunto de dados integrados e formatados compatíveis com o tipo de arquivo aceito pela ferramenta de modelagem. Para realizarmos esta etapa, tivemos que adaptar um formulário *online* semelhante ao questionário socioeconômico utilizado pelo IFPI.

Justificamos a adaptação do questionário para um modelo eletrônico por este ser aplicado aos alunos em formato impresso no momento da matrícula.

De acordo com Marques (2014), para realizarmos trabalhos de análise e MD, é necessário a implantação de um sistema que seja adaptado para captura e armazenagem dos dados. O objetivo principal é deixar os dados no formato favorável ao processo da modelagem.

Com isso, a pesquisa nos forçou a buscar o *Google* Formulário como alternativa de “transcrevermos” os dados impressos para o formato de mídia que fosse compatível com o programa utilizado na etapa da modelagem. Dentre as

funcionalidades do formulário *online*, utilizamos a função exportar, para que a nossa base coletada pudesse ser armazenada numa planilha em formato XLS e CSV (vide **Apêndice A**).

Partindo dessa planilha, inserimos novas colunas intituladas com as notas das disciplinas ofertadas por módulos do Curso Técnico em Segurança do Trabalho. Ao final a planilha foi denominada de “PL_Integrada” e a partir dela construímos o Dicionário Preliminar de Variáveis, conforme as tarefas realizadas e descritas nos **Quadros 4 e 5**, respectivamente.

Quadro 4 - Tarefas da Preparação dos Dados

TAREFA	DESCRIÇÃO
01	Elaboração do formulário <i>online</i> utilizando o <i>Google</i> Formulário com as variáveis do questionário socioeconômico. A criação de um formulário <i>online</i> da plataforma <i>Google</i> , vincula uma planilha eletrônica com dados salvos em nuvem
02	Transcrição dos dados do questionário socioeconômico dos alunos matriculados utilizando o formulário <i>online</i> , obtendo a planilha nomeada como “PL_Cadastro”
03	Conversão do arquivo com resultado final do processo seletivo em formato PDF para XLS utilizando o <i>site</i> www.smallpdf.com obtendo a planilha nomeada como “PL_Seleção”
04	Organização da planilha de mapa de notas conforme a elaboração do PPC do curso. A planilha obtida foi nomeada como “PL_Notas”
05	Integração das planilhas PL_Cadastro, PL_Seleção e PL_Notas para eliminação de duplicidade de dados
06	Obtenção da planilha “PL_Integrada” com dados da amostra, organizados em 199 linhas e 31 colunas
07	Codificação dos nomes das colunas da planilha PL_Integrada e criação do Dicionário Preliminar de Variáveis
08	Classificação das variáveis do dicionário, de acordo com o tipo do dado informado: numérico ou categórico
09	Delimitação e redução das variáveis para a fase da modelagem baseada na literatura sobre o estudo da evasão escolar e MDE

Fonte: Dados da pesquisa (2017).

Quadro 5 - Dicionário Preliminar de Variáveis

VARIÁVEL	TIPO	DESCRIÇÃO
CT	categórico	Cota
PS_P	numérico	Pontuação em Português no processo seletivo de 2013.2 (pontuação máxima: 30)
PS_M	numérico	Pontuação em Matemática no processo seletivo de 2013.2 (pontuação máxima: 30)
TPPS	numérico	Total de pontos no processo seletivo de 2013.2
SX	categórico	Sexo
ID_MAT	numérico	Idade no dia da matrícula (06/09/2013)
EST_C	categórico	Estado civil
CID	categórico	Cidade onde reside
UF	categórico	Estado onde reside

PRO	categórico	Profissão
CID_NAT	categórico	Cidade de nascimento
UF_NAT	categórico	Estado de nascimento
SE_1	categórico	Etnia
SE_2	categórico	Necessidades especiais
SE_2.1	categórico	Tipo de necessidades especiais
SE_3	categórico	Meio de transporte de acesso ao <i>Campus</i> do IFPI
SE_4	categórico	Tipo de escola que o aluno cursou o ensino médio, quanto a fins lucrativos.
SE_5	categórico	Zona da residência do aluno
SE_6	categórico	Grupo de pessoas com os quais o aluno reside
SE_6.1	categórico	Tipo de imóvel onde o aluno reside
SE_7	categórico	Motivo de escolha do curso
SE_8	categórico	Meio de comunicação mais utilizado para se manter informado(a)
SE_9	categórico	Possuir acesso a computador
SE_10	categórico	Tipo de serviço médico quando adoece
SE_11	numérico	Renda familiar
SE_12	numérico	Total de pessoas que residem na casa do aluno, incluindo o aluno e que dependem da renda familiar
SE_13	categórico	Participação do aluno na vida econômica da família
SMDM1	categórico	Situação do aluno com base na média aritmética das disciplinas do módulo 1

Fonte: Dados da pesquisa (2017).

De acordo com Manilla (1994, *apud* Veloso, 2015), a fase da preparação dos dados é o momento que compromete até 80% do processo da mineração em decorrência da necessidade da realização de atividades de pré-processamento tais como: limpeza, integração, transformação, seleção de atributos e redução do volume dos dados.

Com base nessas atividades e precedendo a fase da modelagem, realizamos a exclusão de algumas variáveis citadas no **Quadro 5**, porém conservando o quantitativo de alunos da amostra.

Dentre as variáveis analisadas em nossa pesquisa, foram excluídas CT (Cota), CID_NAT (Cidade de nascimento), UF_NAT (Estado de nascimento), SE_1 (Etnia), SE_2 (Necessidades especiais), SE_2.1 (Tipo de necessidades especiais) e SE_3 (Meio de transporte de acesso ao IFPI).

A variável CT (Cota) foi excluída por termos detectado que o IFPI realiza chamada pública para preenchimento das vagas dos Cursos Técnicos a Distância quando as mesmas não são preenchidas por aprovados ou classificados no processo seletivo para alunos de cursos presenciais e a distância.

Além disso, o questionário socioeconômico utilizado pela Instituição limita-se a coletar dados de alunos selecionados em processo seletivo por sistema de cotas ou ampla concorrência.

As variáveis CID_NAT (Cidade de nascimento), UF_NAT (Estado de nascimento), SE_1 (Etnia), SE_2 (Necessidades especiais) e SE_2.1 (Tipo de necessidades especiais) foram excluídas por serem típicas de estudos que abordam o perfil do aluno evadido em estudos de caso. Porém, até o presente momento, não encontramos sustentação teórica envolvendo essas variáveis em publicações que apontam a relevância, quando estas são combinadas entre si ou com outras variáveis no que tange justificar a evasão escolar. Considerando tratar-se de variáveis que também carregam subjetividades, seria pertinente abrir um espaço reflexivo em torno da questão: até que ponto estes atributos contribuem fidedignamente para a previsão de alunos com tendência a evasão? Ou seria tratar-se de dar ênfase ou evidenciar mais uma questão de caráter excludente/ discriminador?

Por fim, descartamos a variável SE_3 (Meio de transporte de acesso ao *Campus* do IFPI) por entendermos que a mesma é específica de questões direcionadas a alunos de cursos presenciais ofertados em cidades que possuem edificações do IFPI, destoando da modalidade do nosso objeto estudo, uma vez que os encontros presenciais são realizados em polos.

Após essas considerações obtivemos o Dicionário Final de Variáveis, que foram utilizadas na fase da modelagem, explicitadas no **Quadro 6**.

Quadro 6 - Dicionário Final de Variáveis

VARIÁVEL	TIPO	DESCRIÇÃO
PS_P	numérico	Pontuação em Português no processo seletivo de 2013.2 (pontuação máxima: 30)
PS_M	numérico	Pontuação em Matemática no processo seletivo de 2013.2 (pontuação máxima: 30)
TPPS	numérico	Total de pontos no processo seletivo de 2013.2
SX	categórico	Sexo
ID_MAT	numérico	Idade no dia da matrícula (06/09/2013)
EST_C	categórico	Estado civil
CID	categórico	Cidade onde reside
UF	categórico	Estado onde reside
PRO	categórico	Profissão
SE_4	categórico	Tipo de escola que o aluno cursou o ensino médio, quanto a fins lucrativos
SE_5	categórico	Zona da residência do aluno
SE_6	categórico	Grupo de pessoas com os quais o aluno reside
SE_6.1	categórico	Tipo de imóvel onde o aluno reside

SE_7	categórico	Motivo de escolha do curso
SE_8	categórico	Meio de comunicação mais utilizado para se manter informado(a)
SE_9	categórico	Possuir acesso a computador
SE_10	categórico	Tipo de serviço médico quando adoece
SE_11	numérico	Renda familiar
SE_12	numérico	Total de pessoas que residem na casa do aluno, incluindo o aluno e que dependem da renda familiar
SE_13	categórico	Participação do aluno na vida econômica da família
SMDM1	categórico	Situação do aluno com base na média aritmética das disciplinas do módulo 1

Fonte: Dados da pesquisa (2017).

3.2.4 Modelagem

Fase em que um modelo estatístico é explicitado através de uma expressão matemática que combina variáveis independentes, chamadas de explicativas, que podem justificar uma variável dependente ou alvo.

Para esta etapa utilizamos o *software GreatI 2016c*³, visando construir um ou mais modelos utilizando a regressão logística e estimar o nível descritivo de significância (p-valor) de cada variável em conjunto com as demais. É importante lembrar que o p-valor não é fixo e sua variação depende da quantidade de variáveis envolvidas e do quantitativo da amostra.

Além do *software* sobredito utilizamos o *software Excel 2010* na efetuação do cálculo do *score*, visando prognosticar o risco de evasão de cada elemento da amostra, dentro de uma escala de pontuação de valor contínuo de 0 a 1, onde o segundo foi estabelecido como indicador de evasão (alvo).

Justificamos a escolha dos referidos *softwares* por terem sido utilizados na disciplina de Estatística Aplicada do Programa de Pós-Graduação em Tecnologia e Gestão em Educação a Distância, em atividades que envolveram extração de conhecimento em microdados do ENADE 2013, disponíveis para consulta no *site* do INEP. De forma análoga, para realizarmos a etapa de modelagem dos dados da nossa pesquisa, executamos as seguintes tarefas, conforme o **Quadro 7**.

Quadro 7 - Tarefas da Modelagem

TAREFA	DESCRIÇÃO
01	Conversão semântica da planilha PL_Integrada utilizando a função SE no <i>Excel</i> salva no formato XLS e CSV, nomeada como PL_Semântica

³ *Software* livre que compila e interpreta dados econométricos

02	Instalação do <i>software Gretl</i>
03	Importação da planilha PL_Semântica.csv
04	Regressão logística utilizando o algoritmo MQO (Mínimos Quadrados Ordinários) utilizando os dados das planilhas PL_Semântica considerando apenas as variáveis do Quadro 6
05	Transformação das variáveis categóricas em variáveis <i>flag</i> ⁴
06	Aplicação do algoritmo MQO sobre as variáveis <i>flag</i>
07	Identificação das variáveis fortemente preditivas ou explicativas da variável alvo observando o p-valor das variáveis “analisadas” pelo <i>Gretl</i> 2016c: PS_M; CID; PRO; SE_6; SE_6.1; SE_7; SE_8; SE_10; SE_11; SE_12; SE_13 (vide Apêndice G)
08	Identificação das variáveis com maior potencial de justificar a variável alvo observando o p-valor das variáveis “analisadas” pelo <i>Gretl</i> 2016c: PS_M, CID, e SE_11 (vide Apêndice G)
09	Quanto menor o p-valor da variável maior chance de explicar a variável alvo
10	Partição da planilha PL_Semântica em duas outras PL_99_A e PL_99_B, cada uma com 99 registros de estudantes
11	Obtenção do Modelo 1 utilizando o algoritmo MQO (Mínimos Quadrados Ordinários) sobre os dados das planilhas PL_99_A
12	Obtenção do Modelo 2 utilizando o algoritmo MQO (Mínimos Quadrados Ordinários) sobre os dados das planilhas PL_99_B
13	Exportação dos dados em formato XLS, denominada PL_99_A_TRT e PL_99_B_TRT, ambas referentes a dados treinados na modelagem
14	Obtenção de dois conjuntos de treinamentos e dois conjuntos de testes
15	Realização da validação cruzada (<i>cross validation</i>) das duas partições para avaliar a capacidade de generalização dos modelos: <i>2-fold cross</i>
16	Obtenção da planilha PL_ALVO_SCORE
17	Normalização dos <i>scores</i> obtidos
18	Disposição dos <i>scores</i> em rol (na ordem crescente)
19	Análise dos dados da planilha PL_ALVO_SCORE utilizando tabela dinâmica no <i>Excel</i>
20	Representação gráfica dos resultados

Fonte: Dados da pesquisa (2017).

3.2.5 Avaliação

A avaliação é o momento em que podemos identificar os primeiros resultados relacionados com o negócio em estudo. É nesta fase que podemos avaliar o grau de relevância que um ou mais modelos produzidos podem contribuir para o pesquisador alcançar os objetivos definidos na primeira fase da metodologia adotada nesta pesquisa.

A fim de evitarmos redundância de informações sobre os resultados desta pesquisa, utilizamos o Capítulo 4 para concentrarmos as análises e discussões dos resultados obtidos.

⁴Variável que serve para indicar a ocorrência ou não de um fato.

3.2.6 Implantação

A fim de alcançarmos o objetivo geral do presente estudo, não realizamos esta fase. No entanto, a ausência do cumprimento desta etapa não interfere na extração do conhecimento sobre os dados da amostra considerada já processada pelo algoritmo de modelagem.

É importante destacarmos que a fase da Implantação requer tempo para a realização de tarefas que contemplem o plano de distribuição de um modelo compatível com o negócio ou problema em questão, considerando a necessidade de ações no que tange ao monitoramento do desempenho do modelo, manutenção e emissão de um parecer parcial ou final para poder ser distribuído aos clientes (gestores educacionais), comprovando sua relevância.

No entanto, deixaremos como trabalho futuro, a elaboração de um sistema de alerta de risco de evasão aplicável à demanda dos cursos profissionalizantes a distância do IFPI, decorrente da retomada da metodologia CRISP-DM sobre dados coletados com a utilização do produto gerado nesta pesquisa.

4 ANÁLISE E DISCUSSÕES DOS RESULTADOS

Embora este trabalho tenha sido desenvolvido com o objetivo de adaptarmos um processo de MD para estimarmos o risco da evasão escolar em Cursos Técnicos a Distância ofertados pelo IFPI, convém lembrarmos que estimar o risco está diretamente associado às estratégias de prevenção e não de identificação do perfil do aluno já evadido. No entanto, é importante destacarmos que a utilização das técnicas de MD pode ser útil para outras abordagens dentro do contexto educacional, além do fenômeno evasão, como também em estudos que versam sobre a retenção e o sucesso escolar.

Nesse sentido, entendemos que pesquisas que visam analisar o perfil do aluno evadido utilizando MDE podem ampliar o horizonte no que diz respeito à possibilidade da identificação de variáveis nunca ou pouco discutidas na literatura. De fato, é comum encontrarmos estudos que enfatizam atributos como faixa de idade, etnia e gênero sem demonstrar a correlação existente entre esses atributos e o fenômeno evasão. Dessa forma surge o questionamento: até que ponto as análises sobre a evasão, no que remete a faixa etária, etnia e gênero merecem destaque, considerando que a cada dia a sociedade, mesmo tendo sua diversidade, busca a igualdade de direitos e deveres?

Além dessa percepção, é importante que sejam desenvolvidos *softwares* capazes de revelar informações que oportunizem a descoberta do conhecimento, atendendo às necessidades dos profissionais da educação, não se restringindo apenas aos capacitados tecnicamente em inteligência artificial e/ou estatística. Afinal de contas, se a MDE é definida como a descoberta do conhecimento sobre dados provenientes do contexto educacional, então é importante que esta possa ser feita e analisada por um profissional da área da educação: pedagogo, professor, coordenador etc.

Com isso, diante dessas considerações, observamos em nosso estudo que ao utilizarmos o *software Gretl* aplicando o algoritmo MQO sobre a amostra desta pesquisa, foram identificadas as variáveis fortemente explicativas para o alvo evasão: PS_M (Pontuação em Matemática no processo seletivo), CID (Cidade onde reside) e SE_11(Renda familiar).

Assim, ao fazermos uma análise desses dados sob o olhar da literatura com foco na evasão e no relatório do Censo EAD de 2015, construímos algumas hipóteses,

resultante da comparação entre esse referencial teórico e a informação obtida após a modelagem utilizando a nossa base de dados.

Considerando que nossa amostra é constituída por dados de 198 alunos, identificamos o quantitativo de 64 alunos evadidos. Diante destes, ao utilizarmos a ferramenta tabela dinâmica, verificamos que 72% dos evadidos, obtiveram mais de 22 pontos em Matemática no processo seletivo para os Cursos Técnicos a Distância do IFPI. O processo seletivo para o ingresso em cursos EAD do IFPI é composto por provas de Português e Matemática, onde o *score* máximo é de 30 pontos.

Dentre os evadidos, detectamos que 88% residem na mesma cidade polo do curso, além disto, 44% possuem renda familiar de 1 a 2 salários mínimos.

Além destas informações, ao considerarmos a pontuação em Matemática no processo seletivo, percebemos que esta amostra comporta-se de forma diferente das considerações de Araújo e Santos (2012). De acordo com estes autores, alunos com pouca eficiência na educação básica não estão preparados para os conteúdos do ensino profissional.

Neste mesmo sentido, Gaioso (2005), Soares, Ferrão e Marques (2011), defendem que a evasão escolar pode ser motivada pela deficiência no ensino fundamental nas disciplinas de Matemática e Português.

Diante disso, esse percentual nos remete a pensar que estes alunos possivelmente ingressaram em outro curso de nível técnico ou nível superior ligado às ciências exatas, ofertados pelo IFPI ou em outra Instituição.

No que toca às demais variáveis, a nossa amostra apontou informações concordantes ao referido Censo, onde este especifica que a evasão é motivada pelo fator tempo e finanças.

Para Rumberger (2011), as condições socioeconômicas, geralmente medidas pela renda familiar, podem incidir no comportamento do estudante. Além disso, escolas localizadas nos centros urbanos apresentam maior taxa de abandono (RUMBERGER; LIM, 2008).

Quanto ao fator finança, fica difícil construirmos um argumento nos contrapondo a esta argumentação, haja vista que para manter-se em cursos a distância o aluno precisará de dispositivos para acessar a plataforma e um serviço de internet banda larga cujos valores ainda não são acessíveis. Com isso, sinalizamos a importância de estudos sobre quais os impactos e contribuições da estruturação e das condições técnicas dos polos na motivação da permanência de alunos que não

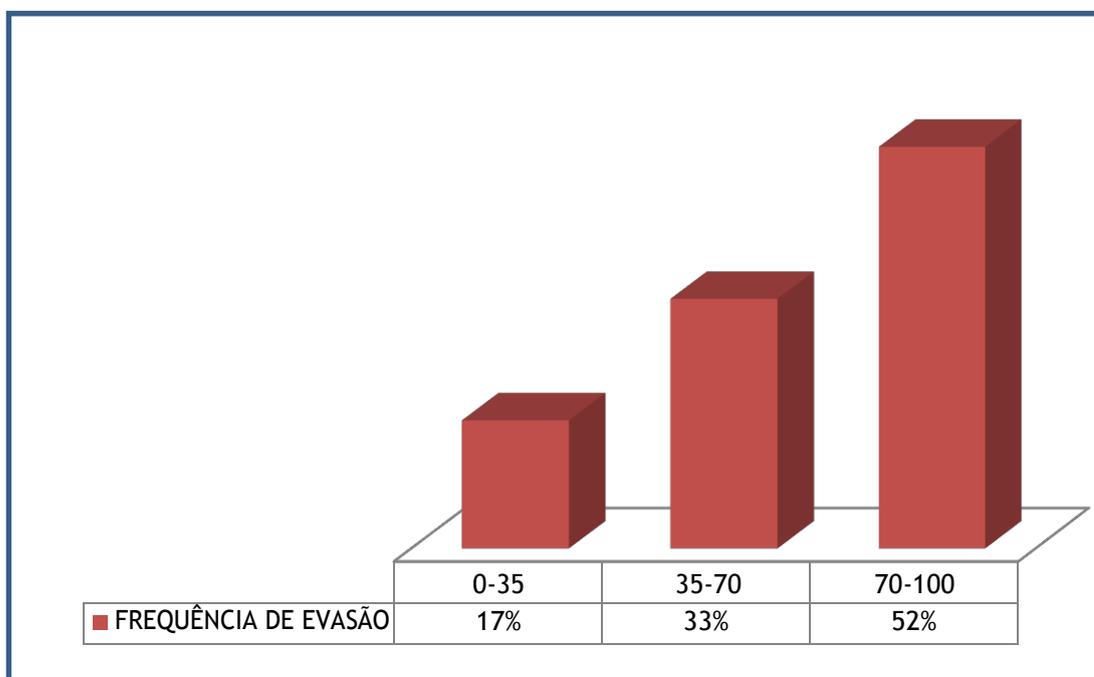
possuem condições financeiras para comprar computador e contratar serviço de *internet* para suas residências.

E ademais, ao analisarmos o fator tempo, entendemos que as Instituições que ofertam cursos a distância devem buscar maneiras de orientar seus discentes no que compete a procedimentos de como organizar estudos e cumpri-los em horários como se estivessem em curso presencial.

Ao darmos prosseguimento à execução da nossa pesquisa com o intuito de analisarmos o risco evasão em Cursos Técnicos a Distância do IFPI, construímos uma tabela composta por duas colunas e 198 linhas. A primeira coluna foi preenchida por 0 ou 1, considerando o valor 1 como situação do aluno evadido, enquanto a segunda coluna era composta pelos *scores* utilizando a equação de Regressão Logística sobre os dados dos alunos ao realizarmos o procedimento de validação cruzada.

Devido à equação supracitada estabelecer o *score* como valor decimal obtido no intervalo fechado de 0 a 1, transformamos esta variação no intervalo fechado de 0 a 100, utilizando a transformação linear denominada normalização. Com isso, obtivemos o **Gráfico 1**.

Gráfico 1 - Distribuição de alunos por faixa de *score*



Fonte: Dados da pesquisa (2017).

Ao visualizarmos o **Gráfico 1**, identificamos que para cada intervalo de pontuação ou *score* está associado um percentual, diretamente proporcional.

Entendemos esse percentual como a intensidade do risco de propensão à evasão que um novo aluno pode praticar, após ser matriculado em um Curso Técnico a Distância do IFPI.

Diante desses valores, identificamos que a média do risco de evasão equivale a 30%, aproximadamente, quando o *score* pertence ao intervalo de 35 a 70 pontos. Baseado nesta conclusão, podemos afirmar que alunos cujo *score* de pontuação for abaixo de 35 pontos, possuem 83% de chance de permanecerem em cursos, conforme nosso objeto de análise.

Enquanto isso, alunos com *scores* maiores que 70 pontos, possuem mais de 50% de probabilidade de escolherem a evasão como solução de seus problemas, antes mesmo de terem o primeiro contato com o curso, após efetuarem suas matrículas.

Portanto, podemos dizer que dados cadastrais de alunos podem sim, serem vistos como fonte de descoberta de conhecimento utilizando técnicas de MD, deixando de servir apenas como dados arquivados, muitas vezes registrados em papel, e vistos como sem utilidade.

Embora a metodologia CRISP-DM possua uma fase destinada à Implantação, deixaremos esta etapa para trabalhos futuros, como dito anteriormente, por necessitarmos de tempo e novas amostras, necessários para a realização de tarefas que possam garantir um modelo ou função que combine variáveis cuja imagem seja de evasão ou permanência.

De maneira análoga ao nosso pensamento, Santos, Siebra e Oliveira (2014) mostraram que é possível propor uma estrutura genérica para prever a evasão de cursos a distância, mesmo com poucos atributos a serem analisados.

Com isso, destacamos que os resultados alcançados em nossa pesquisa não são finalísticos e nem se limitam apenas às variáveis identificadas e analisadas em nosso estudo.

De fato, a aplicação de técnicas de MD para estimarmos o risco de evasão escolar de cursos técnicos a distância do IFPI, não nos impede de acrescentarmos variáveis comportamentais (participação em fóruns, *chats*, notas nas disciplinas, módulos) em trabalhos futuros sobre este tema.

5 CONSIDERAÇÕES FINAIS

Diante deste trabalho podemos concluir que a MD é útil em áreas que possuem dados que nunca ou pouco foram explorados para o planejamento do futuro.

Além disso, entendemos que a MDE utilizando a metodologia CRISP-DM é digna de ser utilizada em diferentes contextos da educação, da mesma forma que a utilizamos para alcançar nosso objetivo geral. Justificamos esta afirmação por percebermos que sua estrutura não obriga ao pesquisador executar todas as etapas, não impõe a técnica a ser adotada, nem tão pouco estabelece qual o algoritmo e *software* serão utilizados para que o conhecimento seja extraído. Ela apenas direciona e permite a tomada de decisão a cada tarefa realizada pelo pesquisador, no sentido de dar prosseguimento ou retornando a etapa anterior.

Neste sentido, destacamos que os estudos sobre a evasão escolar não podem examinar esse fenômeno delimitando-se apenas às variáveis do discente ou tentando argumentar que a análise de um predicado isoladamente comparado com a evasão permita compreender e solucionar o problema já acontecido. É notável a inclusão de muitas outras variáveis a serem investigadas para que possamos identificar a evasão do curso, ou da instituição ou sistema, considerando que o cenário escolar é composto por professores, coordenação, estrutura física e etc.

Contudo, não é simples abraçar todas as variáveis intervenientes em um só estudo, embora seja plausível a inserção e análise de outras dimensões como avaliação da instituição e avaliação dos professores ao término de uma disciplina ou período, possibilitando a tomada de decisão de gestores e planejadores educacionais de forma mais harmônica com a realidade.

A descoberta do conhecimento sobre banco de dados enfatizando a MD, seguindo os moldes da metodologia CRISP-DM, pode motivar pesquisadores e instituições educacionais a construir e implementarem sistemas mais fáceis de serem utilizados por outros atores do cenário escolar, pois os estudos e resultados publicados ainda estão limitados aos profissionais da informática e da estatística.

De fato, despertar o interesse de outros profissionais da educação em conhecer as potencialidades do *Data Mining*, pode reorientar os serviços institucionais no atendimento às especificidades dos estudantes e às exigências do mercado de trabalho.

É importante salientarmos que a cada etapa da metodologia CRISP-DM é possível construirmos parcelas de conhecimento, devido à necessidade de estratégias que favoreçam a realização das tarefas que compõem cada etapa.

Com isso, partindo do quantitativo de publicações com foco na evasão escolar de cursos profissionalizantes a distância e das dificuldades encontradas na etapa da coleta e organização dos dados da amostra, nosso trabalho demandou a criatividade em organizamos nossas leituras e os dados da nossa amostra, explorando as potencialidades do *Google* Formulário na elaboração de sistemas que não dependessem da inserção de um profissional em programação. O *Google* Formulário é um *Google App* que pode ser utilizado por qualquer usuário de *e-mail Gmail*. Sua interface é atraente, intuitiva, de fácil manuseio e acessível para quem desconhece linguagem de programação (vide **Apêndice H**).

Diante das qualidades dessa ferramenta, construímos um sistema de coleta e organização dos dados da amostra adaptado ao questionário socioeconômico aplicado aos alunos de cursos EAD do IFPI, bem como um gerenciador de leituras com detalhes sobre o tipo da publicação, o ano, autores e palavras chave condizentes a nossa pesquisa (vide **Apêndice E**).

Em decorrência da utilização desses sistemas, todos os dados foram armazenados em nuvem, com análise gráfica automática podendo ser exportado em formato XLS, servindo de fonte para a continuidade de nosso trabalho, futuramente.

Portanto, baseado nas leituras, nos dados coletados, dificuldades e possibilidades que surgiram durante a execução da nossa pesquisa, elaboramos um questionário do estudante direcionado ao discentes do EAD do IFPI. Lembramos que este não é um sistema de alerta de risco de evasão, mas um produto elaborado conforme as abordagens dos autores nas referências utilizadas durante nossos estudos sobre a evasão.

Justificamos a elaboração deste formulário por termos percebido que, atualmente, as matrículas dos alunos em Cursos Técnicos a Distância ofertados pelo IFPI ainda utilizam formulários impressos para preenchimento. Reforçamos ainda que as questões do nosso sistema podem nortear a Coordenação Geral do IFPI na busca pelo perfil do aluno matriculado, como também para a identificar alunos evadidos de outras instituições ou de outros cursos ofertados pelo Instituto.

Para ter acesso às questões do referido questionário, o discente deverá possuir uma conta no provedor *Gmail*. Argumentamos esta exigência por ser este provedor

utilizado para criação do *e-mail* institucional dos servidores do IFPI. Ademais, possuir uma conta de *e-mail Gmail* possibilita ampliar novos recursos que podem ser utilizados concomitantemente com a plataforma *Moodle* em atividades de compartilhamento de arquivos, trabalhos colaborativos ou participando de outras salas virtuais como o *Google Classroom*, em momentos de inoperância do ambiente virtual de aprendizagem por motivos de manutenção no *site* do IFPI ou da plataforma.

A imposição de um *e-mail* deste ou de outro provedor no ato da matrícula de um aluno de cursos a distância para ser utilizado durante a sua comunicação com tutores e com a Instituição contribuirá significativamente em atividades que necessitarão obter informações de egressos no que está relacionado com a sua atuação no mercado de trabalho, conforme a sua formação obtida no IFPI.

De fato, segundo Moura (2016), em seu estudo sobre a empregabilidade do egresso do IFPI, a autora afirmou ter dificuldades em localizar os egressos desde as primeiras coletas de dados arquivadas no controle acadêmico da Instituição, principalmente a respeito a *e-mails* inválidos detectados durante sua pesquisa.

Portanto, reforçamos a proposta do nosso produto por considerarmos de grande utilidade para o IFPI e para continuidade de nossos estudos ou de outros pesquisadores, como também por possuir características especiais:

- ✓ É gratuito, então não compromete o orçamento da Instituição;
- ✓ Os dados informados são convertidos em gráficos de forma automática, dispensando a necessidade de um profissional da Informática ou Estatística
- ✓ É editável;
- ✓ Os dados coletados podem ser exportados em formato XLS, CSV, PDF e etc;
- ✓ Os dados coletados ficam salvos em nuvem, não comprometendo o espaço no servidor do IFPI;
- ✓ É compatível com o *e-mail* Institucional dos tutores e coordenadores;
- ✓ Os itens são voltados aos discentes de Cursos Técnicos EAD do IFPI;
- ✓ Subdividido por dimensões conforme a literatura sobre evasão.

A utilização das potencialidades do formulário *online* para efetuarmos a coleta e organização dos dados contribuíram significativamente para a realização das tarefas da fase de modelagem uma vez que o *software Gretl* aceita apenas arquivos em formatos XLS ou CSV. Assim, como aspiramos em trabalhos futuros a criação de um

sistema de alerta de risco de evasão aplicável à demanda dos cursos técnicos a distância do IFPI, nosso produto contribuirá para novas descobertas de conhecimento através da MD de forma mais célere. Desta forma, será possível executarmos as tarefas pertinentes a fase do CRISP-DM chamada Implantação, que não foi executada durante esta pesquisa.

Ademais, em decorrência das tarefas realizadas durante o percurso metodológico sob as etapas da metodologia CRISP-DM, entendemos que independentemente da metodologia adotada neste trabalho, a MD pode ser norteada conforme o passo a passo descrito no **Quadro 8**.

Quadro 8 - Passo a passo da mineração de dados

TAREFAS
Caracterize o negócio: identificação do problema
Fundamente-se teoricamente sobre o problema
Defina objetivos a serem alcançados
Elabore um plano estratégico transformando o negócio em problema de MD
Busque coletar dados que tenham qualidade e relação com o problema do negócio
Formule hipóteses sobre as primeiras impressões dos dados coletados
Organize planilhas em formato XLS e ou CSV com todos os dados acessados
Para dados em mídia de outros formatos, será necessária a utilização de conversores de arquivos disponíveis na <i>web</i>
Para dados em formato impresso, convém transcreve-los utilizando um formulário online com exportação em formato XLS e ou CSV
Delimite as variáveis a serem utilizadas: integre, transforme e reduza os dados
Integre os dados em um única planilha
Crie um dicionário preliminar de variáveis
Codifique e classifique as variáveis em numéricas ou categóricas
Crie um dicionário final de variáveis
Retome a planilha que possui os dados integrados
Busque um <i>software</i> de mineração para a realização da modelagem
Utilize um <i>software</i> que construa um ou mais modelos utilizando a regressão logística e estime o nível de significância (p-valor) de cada variável
Transforme as variáveis categóricas em variáveis que possam indicar a ocorrência ou não de um fato (<i>flag</i>)
Identifique as variáveis com maior potencial de justificar a variável alvo
Particione a planilha em duas iguais de dados com as mesmas variáveis: planilhas A e B
Aplique o algoritmo de modelagem nas duas planilhas para obter dois modelos
Realize a validação cruzada das duas partições
Construa uma planilha com dados contendo o alvo e <i>score</i>
Utilize as funções matemática da planilha eletrônica para efetuar o cálculo do <i>score</i> e escolha o valor a ser indicado para ser o alvo (0 ou 1)
Normalize os <i>scores</i> obtidos e depois os coloque em rol (na ordem crescente)
Utilize a função tabela dinâmica para analisar os resultados
Represente graficamente os resultados
Avalie os resultados comparando-os às suas hipóteses, objetivos e referencial teórico
Escolha um dos modelos obtidos e aplique-o em uma nova base de dados
Avalie novamente os resultados
Crie um sistema que melhor se ajuste ao modelo a ser utilizado
O sistema deverá ter interface amigável ao usuário final
Avalie os resultados e a necessidade de inserção ou exclusão de variáveis

Fonte: Dados da pesquisa (2017).

Observando o **Quadro 8** percebemos que, em geral as tarefas a serem realizadas para a descoberta do conhecimento sobre dados de grande volume, independem do contexto do problema em questão. No entanto, estas tarefas dependem principalmente dos atributos que melhor expressam as características de um grupo e da relação existente entre o estudo e a seleção do algoritmo acertado para solução do problema.

Com isso, retornando ao objetivo geral e às análises dos resultados deste trabalho, entendemos que a MDE pode contribuir não somente para tomada de decisão dos gestores educacionais na predição da evasão escolar partindo de características socioeconômicas anteriores ao ingresso do aluno, mas do desempenho acadêmico durante o curso, como também na abordagem investigativa para outro problema educacional como a retenção escolar, ou no sucesso da diplomação.

Portanto, esperamos que os gestores dos cursos a distância do IFPI bem como de outras Instituições, possam valer-se deste estudo e do produto desta pesquisa para a elaboração de uma política educacional voltada as suas realidades promovendo o envolvimento e a responsabilidade dos alunos na construção de uma sociedade melhor.

REFERÊNCIAS

- ABED. Censo EaD.BR 2015: **relatório analítico de aprendizagem a distância no Brasil**. Curitiba: InterSaberes, 2016.
- ADEODATO, P. J. L.; SANTOS FILHO, M. M.; RODRIGUES, R. L. Predição de desempenho de escolas privadas usando o ENEM como indicador de qualidade escolar. **Anais do Simpósio Brasileiro de Informática na Educação**. v. 25. n. 1, 2014.
- ALVES-MAZZOTTI, A. J.; GEWANDSZNAJDER, F. **O método nas ciências naturais e sociais**. São Paulo: Pioneira Thomson Learning, 2004.
- APPOLINÁRIO, F. **Dicionário de Metodologia Científica**. 2. ed. São Paulo: Atlas, 2011.
- ARAÚJO, C. F.; SANTOS R. A. **A Educação Profissional de Nível Médio e os Fatores Internos / Externos às Instituições que Causam a Evasão Escolar**. The 4th International Congress on University-Industry Cooperation – Taubate, SP – Brazil – December 5th through 7th, 2012.
- BAKER, R.; ISOTANI, S.; CARVALHO, A. 2011. Mineração de dados educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**. v 19. n 2., 2011.
- BARBOSA, W.; MÁXIMO, D.; JATOBÁ, A.; LEITE, A.; SOARES, E. . **Uma Proposta para identificação de causas da evasão na educação a distância através de mineração de dados**. 2014. Disponível em: <erbase2014.uefs.br/artigos/125801.pdf>. Acesso em: 18 mar. 2016.
- BARRETO, A. A. A eficiência técnica e econômica e a viabilidade de produtos e serviços de informação. **Ciência da Informação**, v. 25, n. 3, p. 1-18, 1996. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/466/425>> p.1-18> Acesso em: 10 set. 2016.
- CAMILO, C. O.; SILVA, J. C. da. **Mineração de Dados**: conceitos, tarefas, métodos e ferramentas. Instituto de Informática da Universidade Federal de Goiás, 2009.
- CARDOSO, O. N. P.; MACHADO, R. T. M. **Gestão do conhecimento usando data mining**: estudo de caso na Universidade Federal de Lavras. 2008. Disponível em: <<http://www.scielo.br/pdf/rap/v42n3/a04v42n3.pdf>>. Acesso em: 20 abr. 2016.
- CARVALHO, M. P. Sucesso e fracasso escolar: uma questão de gênero. **Educação e Pesquisa**, São Paulo, v.29, n.1, p. 185-193, jan./jun. 2003.
- CASTRO, M. H. G.; TIEZZI, S. Os desafios da educação no Brasil: a reforma do ensino médio e a implantação de Enem no Brasil. In: BROCK, Colin; SCHWARTZMAN, Simon. **Os desafios da educação no Brasil**. Rio de Janeiro:

Nova Fronteira, 2005, pp. 119-152.

CHAPMAN, P. **CRISP-DM 1.0: Step-By-Step Data Mining Guide**. [S.I.]: 2000. Disponível em: <<http://www.crisp-dm.org/download.htm>>. Acesso em: 28 jan. 2015.

CONFEDERAÇÃO NACIONAL DA INDÚSTRIA. **Departamento Nacional, Mapa Estratégico da Indústria 2013-2022**. Brasília, CNI, 2013. Disponível em:<www.portaldaindustria.org.br>. Acesso em: 30 out 2016.

COSTA, E.; BAKER, R. S. J. D.; AMORIM, L.; MARINHO, J. M. T. **Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações**. Jornada de Atualização em Informática na Educação - JAIE 2012.

DATE, C. J. **An Introduction to Databases Systems**. 8. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2003.

DENZIN, N. K. & LINCOLN, Y. S. **Handbook of Qualitative Research**. Thousand Oaks: Sage, 2005.

DORE, R.; LÜSCHER, A. Z. **Permanência e evasão na educação técnica de nível médio em Minas Gerais**. Cad. Pesqui., São Paulo, v. 41, n. 144, 2011.

FAYYAD, U.; SHAPIRO, G. P.; SMYTH, P. **From Data Mining to Knowledge Discovery: An Overview**. Menlo Park, CA: AAAI Press/The MIT Press, 1996.

FIGUEIREDO, N. G. S.; SALLES, D. M. R. **Educação Profissional e evasão escolar em contexto: motivos e reflexões**. Disponível em:<<http://dx.doi.org/10.1590/s0104-40362017002500397>>. Acesso em: 25 jul. 2016.

FONSECA,, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002.

FREIXO, M. **Metodologia Científica: fundamentos, métodos e técnicas**. Lisboa: Instituto Piaget, 2011.

GAIOSO, N. P. L. **Evasão discente na educação superior: a perspectiva dos dirigentes e dos alunos**. 2005. 99 f. Dissertação (Mestrado em Educação) – Universidade Católica de Brasília, Brasília, 2005.

GARCIA, E.; ROMERO, C.; VENTURA, S.; CASTRO, C. A collaborative educational association rule mining tool. **The Internet and Higher Education**. v. 14 n. 2, 2011.

GIL, A. C. **Projeto de Pesquisa**. 3. ed. São Paulo: Atlas, 1996.

JOHANN, C. C. **Evasão escolar no Instituto Federal Sul-Rio-Grandense: um estudo de caso no Campus Passo Fundo**. 2012. 119 f. Dissertação (Mestrado em Educação) - Universidade de Passo Fundo, Passo Fundo, 2012.

KDNUGGETS. Disponível em: < <http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html> >. Acesso em: 3 de julho de 2016.

LAROSE, D. T. **Discovering Knowledge in Data: An Introduction to Data Mining**. 2005. Disponível em:<<http://onlinelibrary.wiley.com/book/10.1002/0471687545>>. Acesso em: 17 nov. 2016.

LOBO, M. B. C.M. **Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções**. 2012. Disponível em:<http://www.institutolobo.org.br/imagens/pdf/artigos/art_087.pdf>. Acesso em: 20 jan. 2017.

MACHADO, R. D.; NARA, E. O. B.; SCHREIBER, J. N. C.; SCHWINGEL, G. A. Estudo bibliométrico em mineração de dados e evasão escolar. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 11., 2015, Rio de Janeiro. **Anais...**, Rio de Janeiro: INOVARSE, 2015.

MARQUES, J. L. Q. **Mineração de Dados Educacionais: um estudo de caso utilizando o Ambiente Virtual do SENAI**. 2014. 72 f. Monografia (Licenciatura Plena em Computação) - Universidade Estadual da Paraíba, Campina Grande, 2014.

MIRANDA, R. C. da R. O uso da informação na formulação de ações estratégicas pelas empresas. **Ci. Inf.**, Brasília, v. 28, n. 3, 1999. Disponível em:<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651999000300006&lng=pt&nrm=iso>. Acesso em: 30 maio 2016.

MORAES, J. O.; THEÓPHILO, C. R. **Evasão no Ensino Superior: Estudo dos fatores causadores da evasão no Curso de Ciências Contábeis da Universidade Estadual de Montes Claros**. 2010. Disponível em:<<http://www.congressosp.fipecafi.org/anais/artigos102010/419.pdf>>. Acesso em: 15 set. 2016.

MOURA, J. M. M. O. **A efetividade da educação a distância na visão do egresso: uma análise sobre a empregabilidade**. 2016. 171f. Dissertação (Mestrado em Tecnologia e Gestão em Educação a Distância) – Universidade Federal Rural de Pernambuco, Recife, 2016.

NAVEGA, S. Princípios essenciais do data minig. In: INFOIMAGEM. 2002. **Anais. Cenadem**, nov. 2002. Disponível em : <www.intelliwise.com/snavega>. Acesso em: 28 jun. 2016.

OLIVEIRA, S. L. **Tratado de Metodologia Científica**. São Paulo: Editora Pioneira, 1998.

PORTAL BRASIL. **MEC cria grupo para examinar causa de evasão escolar**. Disponível em:< <http://www.brasil.gov.br/educacao/2013/11/mec-cria-grupo-para-examinar-causa-de-evacao-escolar> >. Acesso em: 30 maio 2016.

PORTAL TODOS PELA EDUCAÇÃO. **Brasil tem 3ª maior taxa de evasão escolar entre 100 países, diz Pnud**, 2013. Disponível em: <<http://www.todospelaeducacao.org.br/educacao-na-midia/indice/26226/brasil-tem-3-maior-taxa-de-evasao-escolar-entre-100-paises-diz-pnud/>>. Acesso em: 30 maio 2016.

REIS, F. L. **Como elaborar uma dissertação de mestrado**. Lisboa: Pactor, 2010.

RODRIGUEZ, A. **Fatores de permanência e evasão de estudantes no ensino superior privado brasileiro: um estudo de caso**. 2011. Disponível em: <<http://revistas.pucsp.br/index.php/caadm/article/view/9009>>. Acesso em: 30 mar. 2016.

RUMBERGER, R. ; LIM, S. **Why Students Drop Out of School: A Review of 25 Years of Research**, California Dropout Research Project, Santa Barbara, 2008.

RUMBERGER, R. **Dropping out: Why students drop out of high school and what can be done about it**. Cambridge: Harvard University Press, 2011.

SANTOS, R. N.; SIEBRA, C. A.; OLIVEIRA, E. D. S. **Uma abordagem genérica de identificação precoce de estudantes com risco de evasão em um AVA utilizando técnicas de mineração de dados**. 2014. Disponível em: <<http://docplayer.com.br/712273-Uma-abordagem-generica-de-identificacao-precoce-de-estudantes-com-risco-de-evasao-em-um-ava-utilizando-tecnicas-de-mineracao-de-dados.html>>. Acesso em: 30 mar. 2016

SILVA FILHO, R. L. L.; LOBO, M. B. de C. M.; HIPÓLITO, O. **Evasão no Ensino Superior: causas e remédios**. 2009. Disponível em: <<http://robertolobo.com.br/index.php/2009/06/evasao-no-ensino-superior-causas-e-remedios/>>. Acesso em: 30 de maio 2016.

SILVA FILHO, R. L. L.; MOTEJUNAS, P. R.; HIPÓLITO, O., LOBO, M. B. C. M. **A Evasão no Ensino Superior Brasileiro**. 2007. Disponível em: <<http://dx.doi.org/10.1590/S0100-15742007000300007>>. Acesso em: 30 ago. 2016.

SILVA, C. R. **Evasão nos cursos de educação a distância: o caso do campus Floriano do Instituto Federal do Piauí**. 2016. 90f. Dissertação (Mestrado em Tecnologia e Gestão em Educação a Distância) – Universidade Federal Rural de Pernambuco, Recife, 2016.

SILVA, E. L. da. **Metodologia da pesquisa e elaboração de dissertação**. 4 ed. Atual. Florianópolis: UFSC, 2005.

SILVA, R. E. D.; RAMOS, J. L. C.; RODRIGUES, R. L.; GOMES, A. S.; FONSÊCA, J. A. V. Mineração de dados educacionais na análise das interações dos alunos em um Ambiente Virtual de Aprendizagem. **Anais do Simpósio Brasileiro de Informática na Educação**. v. 26. n. 1., 2015.

SOARES, M. T.; FERRÃO, E. M; MARQUES A. C. **Análise da evasão no ProJovem Urbano**: uma abordagem através do Modelo de Regressão Logística Multinível. Aval. Pol. Públ. Educ., Rio de Janeiro, v. 19, n. 73, p. 841-860, out./dez. 2011.

VALENTIM, M. L. P. Inteligência competitiva em organizações: dado, informação e conhecimento. **Data Grama Zero**, Rio de Janeiro, v.3., n.4, p.1-13, ago. 2002. Disponível em:<http://www.dgz.org.br/ago02/F_I_art.htm>. Acesso em: 25 ago. 2016.

VELOSO, L. A. **A predição da evasão escolar dos cursos técnicos de nível médio**: um estudo de caso no SENAI. 2015.94f. Dissertação (Mestrado em Gestão do Conhecimento e da Tecnologia da Informação) – Universidade Católica de Brasília, Brasília, 2015.

WAZLAWICK, R. S. **Metodologia de Pesquisa para Ciência da Computação**. Rio de Janeiro: Elsevier, 2008.

ZORDAN, G. R. **Estudo sobre os fatores que influenciam a evasão no curso de Administração da UFMA, na modalidade a distância, no polo de apoio presencial de Porto Franco - MA**. 2012. 103 f. Dissertação (Programa de Pós-Graduação Interdisciplinar e Cultura e Sociedade, da Universidade Federal do Maranhão) – Universidade Federal do Maranhão, Maranhão, 2012.

APÊNDICE A – COLETA DOS DADOS



Fonte: Dados da pesquisa (2017).

APÊNDICE B – TERMO DE AUTORIZAÇÃO DE USO DE IMAGEM

TERMO DE AUTORIZAÇÃO DE USO DE IMAGEM ADULTO

Eu, **Andrea Larisse Meneses Rodrigues**, 29 anos, nacionalidade **brasileira**, estado civil **solteira**, portador da Cédula de identidade RG nº. **5.032.540-SSP-PL**, inscrito no CPF/MF sob nº **042.065.423-25**, residente à **Rua dos Aflitos**, nº. **270**, Bairro **Piçarra**, município de **Teresina**, Estado **Piauí**. AUTORIZO o uso de minha imagem em publicações da dissertação Predição do risco de evasão dos Cursos Técnicos a distância do IFPI: uma aplicação da metodologia CRISP-DM em dados educacionais, do Programa de Pós Graduação: Mestrado Profissional em Tecnologia e Gestão em Educação a Distância UFRPE. A presente autorização é concedida a título gratuito, abrangendo o uso da imagem acima mencionada em todo território nacional e no exterior, das seguintes formas: (I) *out-door*; (II) *busdoor*; folhetos em geral (encartes, mala direta, catálogo, etc.); (III) folder de apresentação; (IV) anúncios em revistas e jornais em geral; (V) home page; (VI) cartazes; (VII) back-light; (VIII) mídia eletrônica (painéis, vídeo-tapes, televisão, cinema, programa para rádio, entre outros). Por esta ser a expressão da minha vontade declaro que autorizo o uso acima descrito sem que nada haja a ser reclamado a título de direitos conexos à minha imagem ou a qualquer outro, e assino a presente autorização em 02 vias de igual teor e forma.

Teresina, dia **15** de **Outubro** de **2016**.

Andrea Larisse Meneses Rodrigues

(assinatura)

Telefone p/ contato: 86-98151-2464

APÊNDICE C – FORMULÁRIO ONLINE DA COLETA DOS DADOS

COLETA E ORGANIZAÇÃO DOS DADOS

NOME
Sua resposta: _____

NOME SOCIAL
Sua resposta: _____

COTA
 SIM
 NÃO

CARACTERIZAÇÃO DO CURSO

NOME DO CURSO *
 SEGURANÇA DO TRABALHO - A
 SEGURANÇA DO TRABALHO - B
 SEGURANÇA DO TRABALHO - ZSUL - PALMEIRAS
 EVENTOS - ZSUL - PALMEIRAS

ANO LETIVO *
 2013.2

MATRÍCULA
Sua resposta: _____

CPF *
Sua resposta: _____

SEXO
 M
 F

DATA DE NASCIMENTO
 Dia: _____
 Mês: _____
 Ano: _____

ESTADO CIVIL
 SOLTEIRO(A)
 CASADO(A)
 SEPARADO(A)
 VIÚVO(A)

CIDADE
Sua resposta: _____

UF
Sua resposta: _____

PROFISSÃO
Sua resposta: _____

CIDADE NATAL
Sua resposta: _____

UF NATAL
Sua resposta: _____

QUESTIONÁRIO SÓCIO ECONÔMICO EDUCACIONAL / TÉCNICO 2013.2

1) Você se considera:
 Branco(a)
 Negro(a)
 Indígena
 Pardos(a)
 Amarelo(a)

2) Você tem necessidades especiais?
 Sim
 Não

10) Quando você adoece recorre a quais serviços médicos?
 Hospital público / SUS
 Plano de saúde
 Serviço médico particular
 Outro: _____

11) Qual a renda mensal da sua família?
 Até 1 salário
 1 a 2 salários
 2 a 3 salários
 3 a 5 salários
 5 a 10 salários
 10 a 20 salários
 Mais de 20 salários

12) Total de pessoas que residem na casa (incluindo você) que dependem da renda familiar.
 1 pessoa
 2 pessoas
 3 pessoas
 4 pessoas
 5 pessoas
 6 pessoas
 Acima de 6 pessoas

13) Qual sua participação na vida econômica da sua família?
 Não trabalha e é sustentado pela família ou por outras pessoas.
 Trabalha, mas recebe ajuda financeira da família ou de outras pessoas.
 Trabalha e é responsável pelo sustento, além de contribuir parcialmente para o sustento da família.
 Trabalha e é responsável apenas pelo seu próprio sustento.
 Trabalha e é o principal responsável pelo sustento da família.

ENTAR

Mostrar esse formulário para todos os Google

Este formulário foi criado em INSTITUTO FEDERAL DO PIAUÍ - Demandas Alunos - Tópicos de Serviço - Tópicos Alunos

Google Formulários

2.1) Em caso afirmativo, indique o tipo
 Físico
 Intelectual
 Auditiva
 Visual
 Outro: _____

3) Meio de transporte utilizado para vir ao IFPI?
 A pé
 Ônibus
 Carro ou moto
 Bicicleta
 Metrô

4) A escola onde você cursa/curso ou Ensino Médio é:
 Pública
 Privada
 Filantrópica

5) Local de sua residência
 Zona
 Urbana
 Rural

6) Atualmente, você reside:
 Com os pais
 Com parentes
 Com amigos
 Casa do estudante
 Solteiro

6.1) Sua casa é:
 Própria
 Alugada
 Outro: _____

7) Qual o principal motivo de sua escolha pelo curso?
 Credibilidade na instituição
 Opção por uma formação técnica
 Identificação com a área/ curso escolhido
 Influência da família e/ou colegas
 Só o IFPI oferece o curso que você quer
 Outro: _____

8) Qual é o meio de comunicação que você mais utiliza para se manter informado?
 Livros
 TV
 Rádio
 Internet
 Outro: _____

9) Você tem acesso ao computador?
 Sim, para lazer e trabalhos escolares
 Sim, para trabalhos profissionais
 Sim, para outros fins
 Não

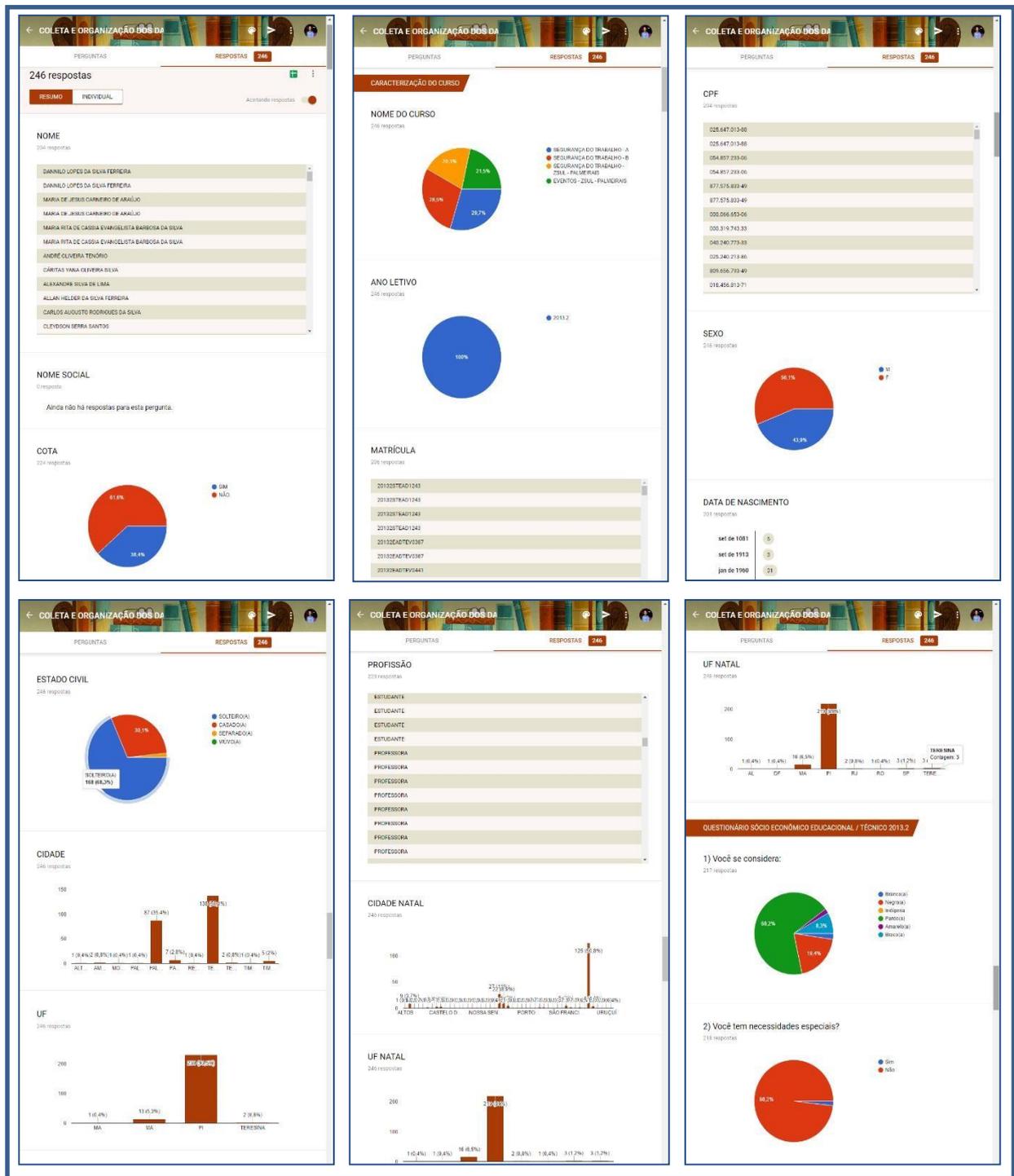
10) Quando você adoece recorre a quais serviços médicos?
 Hospital público / SUS
 Plano de saúde
 Serviço médico particular
 Outro: _____

11) Qual a renda mensal da sua família?
 Até 1 salário
 1 a 2 salários
 2 a 3 salários
 3 a 5 salários
 5 a 10 salários
 10 a 20 salários
 Mais de 20 salários

12) Total de pessoas que residem na casa (incluindo você) que dependem da renda familiar.
 1 pessoa
 2 pessoas
 3 pessoas
 4 pessoas
 5 pessoas
 6 pessoas
 Acima de 6 pessoas

Fonte: Dados da pesquisa (2017).

APÊNDICE D – RESUMO DOS DADOS COLETADOS



Fonte: Dados da pesquisa (2017).



Fonte: Dados da pesquisa (2017).

APÊNDICE E – GERENCIAMENTO DAS LEITURAS

GERENCIADOR DE LEITURAS - MINERAÇÃO DE DADOS EDUCACIONAIS

Obrigado

ANO DE PUBLICAÇÃO *

S/A

Outro

TIPO DE PUBLICAÇÃO *

ARTIGO

DISSERTAÇÃO

MANUAL

MONOGRAFIA

SITE

TESE

VÍDEO

Outro

TÍTULO *

Sua resposta

AUTORES *

Sua resposta

LINK *

Sua resposta

PALAVRA CHAVE 01 *

EDUCAÇÃO

MINERAÇÃO DE DADOS

MINERAÇÃO DE DADOS EDUCACIONAIS

Outro

PALAVRA CHAVE 02 *

EAD

PRESENCIAL

Outro

PALAVRA CHAVE 03 *

FUNDAMENTAL

MÉDIO

TÉCNICO

SUPERIOR

Outro

PALAVRA CHAVE 04 *

EVASÃO

Outro

PALAVRA CHAVE 05 *

PREDIÇÃO

PERFIL

ESTUDO DE CASO

Outro

PALAVRA CHAVE 06 *

CRISP-DM

Outro

PALAVRA CHAVE 07 *

Sua resposta

Enviar

Página 1 de 1

Este formulário foi criado em INSTITUTO FEDERAL DO PIAUÍ - Secretaria de Apoio - Serviço de Apoio - Formas Adicionais

Google Formulários

Fonte: Dados da pesquisa (2017).

APÊNDICE G – VARIÁVEIS FORTEMENTE PREDITIVAS

Modelo 1: MQO, usando as observações 1-198
Variável dependente: SMDM1

	<i>Coefficiente</i>	<i>Erro Padrão</i>	<i>razão-t</i>	<i>p-valor</i>	
const	-0,64018	0,274043	-2,3361	0,0206	**
PS_M	0,0348084	0,00993351	3,5041	0,0006	***
DCID_1	-0,171895	0,0821269	-2,0930	0,0377	**
DPRO_1	-0,088865	0,0725317	-1,2252	0,2221	
DSE_61_1	0,114601	0,0914858	1,2527	0,2119	
DSE_7_2	0,118204	0,0702715	1,6821	0,0943	*
DSE_7_3	0,171356	0,0965958	1,7739	0,0777	*
DSE_8_3	0,186712	0,185318	1,0075	0,3150	
DSE_10_1	-0,0975084	0,0727016	-1,3412	0,1815	
DSE_11_3	0,157585	0,0689005	2,2871	0,0233	**
DSE_12_1	0,176295	0,0988231	1,7839	0,0761	*
DSE_12_2	0,12559	0,104747	1,1990	0,2321	
DSE_13_1	-0,162834	0,08696	-1,8725	0,0627	*
DSE_13_2	-0,105299	0,0956447	-1,1009	0,2724	
DSE_13_3	-0,121621	0,086433	-1,4071	0,1611	
Média var. dependente	0,323232	D.P. var. dependente	0,468896		
Soma resíd. quadrados	35,25986	E.P. da regressão	0,438950		
R-quadrado	0,185931	R-quadrado ajustado	0,123653		
F(14, 183)	2,985485	P-valor(F)	0,000385		
Log da verossimilhança	-110,1232	Critério de Akaike	250,2463		
Critério de Schwarz	299,5703	Critério Hannan-Quinn	270,2110		

Fonte: Dados da pesquisa (2017).

APÊNDICE H – PRODUTO DA PESQUISA

MINISTÉRIO D
SECRETARIA DE EDUCAÇÃO CIENTÍFICA E T
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
PRÓ-RETORIA
DIRETORIA

QUESTIONÁRIO DO(A) ESTUDANTE
EAD IFPI

Careja estudante.
Este questionário constitui um instrumento extremamente importante para compor o perfil socioeconômico e tecnológico dos alunos dos Cursos Técnicos à Distância do IFPI. Sua confidencialidade é garantida e a grande expectativa é para melhor conhecimento e desempenho dentro do Sistema à Distância de qualidade e excelência. As respostas às questões serão analisadas em conjunto, preservando o sigilo de sua identidade. Para responder, basta clicar sobre a alternativa desejada. O sistema gravará a resposta no banco de dados do IFPI. O questionário será enviado ao IFPI apenas quando, na última página, for acionado o botão "Enviar", indicando o preenchimento total do questionário. Ao clicar no botão "Cancelar", será possível "EDITAR" suas respostas imediatamente. Lembremos que ao fechar a página ou seu navegador, não será mais possível acessar os seus dados informados no formulário. Portanto, faça ou equivoque após enviar suas respostas, recomentando o informe ao coordenador do Polo ou à Coordenação para que possa preencher um novo Questionário.
Agradecemos a sua colaboração!

*Obrigado

Endereço de e-mail *

Seu e-mail

1. CARACTERIZAÇÃO PRELIMINAR DO DISCENTE

1.1 Qual o pólo do seu curso? *

Pólo 01
 Pólo 02
 Pólo 03

1.2 Qual o seu curso? *

Curso Técnico X
 Curso Técnico Y
 Curso Técnico Z

1.3 Qual módulo você irá cursar? *

I
 II
 III
 IV

1.4 Nome Completo *

1.4. Nome Completo *

Sua resposta

1.5. CPF *

Sua resposta

1.6. Assinale a alternativa que representa a forma como ingressou neste curso. *

Cota
 Ampla concorrência
 Chamada pública

2. DIMENSÃO SOCIAL

2.1 Qual o seu estado civil? OBS: Para os casos de separado, separado(a) judicialmente/divorciado(a) e viúvo(a), assinale a alternativa "Solteiro(a)". *

Solteiro(a)
 Casado(a)

2.2 Qual sua idade? *

Sua resposta

2.3 Quanto a sua etnia, como você se considera? *

Branco(a)
 Negro(a)
 Pardo(a)/mulato(a)
 Amarelo(a)
 Indígena

2.4. Qual a sua nacionalidade? *

Brasileira
 Brasileira naturalizada
 Estrangeira

2.5. Você possui alguma necessidade especial? *

Sim
 Não

2.6. Se a resposta da pergunta anterior for "Sim", informe o tipo

2.6. Se a resposta da pergunta anterior for "Sim", informe o tipo de necessidade na opção "Outro". Caso contrário assinale "Não possui". *

Não possui
 Outro

3. DIMENSÃO FAMILIAR (background familiar)

3.1. Com quem você mora atualmente? *

Sozinho(a)
 Com pais
 Com parentes
 Com cônjuge e/ou filhos
 Amigos(as)
 Outro:

3.2. Quantas pessoas moram com você? *

Nenhuma
 Uma pessoa
 Duas pessoas
 Três pessoas
 Quatro pessoas
 Outro:

3.3. Qual etapa da escolarização seu pai concluiu? *

Nenhuma
 Ensino fundamental: 1º ao 5º ano (1ª a 4ª série)
 Ensino fundamental: 6º ao 9º ano (5ª a 8ª série)
 Ensino médio
 Educação superior
 Pós-graduação

3.4. Qual etapa da escolarização sua mãe concluiu? *

Nenhuma
 Ensino fundamental: 1º ao 5º ano (1ª a 4ª série)
 Ensino fundamental: 6º ao 9º ano (5ª a 8ª série)
 Ensino médio
 Educação superior
 Pós-graduação

4. DIMENSÃO ECONÔMICA

4.1 Você trabalha? *

Sim
 Não

4.2. Sua moradia é: *

Própria
 Alugada
 Cediada

4.3. Quando você adoecer, a qual sistema de saúde você recorre? *

SUS
 Plano de saúde
 Médico particular
 Outro:

4.4. Qual sua renda individual? *

Nenhuma, pois não trabalho.
 Menos de um salário mínimo.
 Apenas um salário mínimo.
 De 2 a 4 salários mínimos.
 Acima de 4 salários mínimos.

4.5. Qual a renda total de sua família? (Incluindo você) *

Menos de um salário mínimo.
 Apenas um salário mínimo.
 De 2 a 4 salários mínimos.
 Acima de 4 salários mínimos.

4.6. Quantas pessoas, incluindo você, dependem da renda familiar? *

Nenhuma, apenas eu, pois moro sozinho(a).
 Duas pessoas
 De duas a quatro pessoas.
 Mais de quatro pessoas.

4.7. Qual sua participação na economia da sua família? *

Nenhuma, pois não trabalho.

Fonte: Dados da pesquisa (2017).

<p>4.7. Qual sua participação na economia da sua família? *</p> <p><input type="radio"/> Nenhuma, pois não trabalho.</p> <p><input type="radio"/> Nenhuma, pois meu salário é apenas para minhas despesas.</p> <p><input type="radio"/> Contribui parcialmente para o sustento da minha família.</p> <p><input type="radio"/> Contribui totalmente para o sustento da minha família.</p> <p>5. DIMENSÃO EDUCACIONAL</p> <p>5.1. Em que tipo de escola você cursou o Ensino Fundamental? *</p> <p><input type="radio"/> Todo em escola pública.</p> <p><input type="radio"/> Todo em escola privada (particular).</p> <p><input type="radio"/> A maior parte em escola pública.</p> <p><input type="radio"/> A maior parte em escola privada (particular).</p> <p>5.2. Em que tipo de escola você cursou o Ensino Médio? *</p> <p><input type="radio"/> Todo em escola pública.</p> <p><input type="radio"/> Todo em escola privada (particular).</p> <p><input type="radio"/> A maior parte em escola pública.</p> <p><input type="radio"/> A maior parte em escola privada (particular).</p> <p>5.3. Que tipo de Ensino Médio você concluiu? *</p> <p><input type="radio"/> Ensino médio tradicional.</p> <p><input type="radio"/> Profissionalizante técnico integrado ao médio.</p> <p><input type="radio"/> Profissionalizante magistério (Curso Normal).</p> <p><input type="radio"/> Educação de Jovens e Adultos (EJA) ou Supletivo.</p> <p><input type="radio"/> Outro: _____</p> <p>5.4. Durante o Ensino Fundamental I: 1º ao 5º Ano (1ª a 4ª série), você foi "Reprovado(a)" em: *</p> <p><input type="radio"/> Matemática.</p> <p><input type="radio"/> Língua Portuguesa.</p> <p><input type="radio"/> Matemática e Língua Portuguesa.</p> <p><input type="radio"/> Outra(s) disciplinas.</p> <p><input type="radio"/> Nenhuma disciplina.</p> <p>5.5. Durante o Ensino Fundamental II: 6º ao 9º Ano (6ª a 8ª série), você foi "Reprovado(a)" em: *</p> <p><input type="radio"/> Matemática.</p> <p><input type="radio"/> Língua Portuguesa.</p>	<p>5.5. Durante o Ensino Fundamental II: 6º ao 9º Ano (6ª a 8ª série), você foi "Reprovado(a)" em: *</p> <p><input type="radio"/> Matemática.</p> <p><input type="radio"/> Língua Portuguesa.</p> <p><input type="radio"/> Matemática e Língua Portuguesa.</p> <p><input type="radio"/> Outra(s) disciplinas.</p> <p><input type="radio"/> Nenhuma disciplina.</p> <p>5.6. Durante o Ensino Médio, você foi "Reprovado(a)" em: *</p> <p><input type="radio"/> Matemática.</p> <p><input type="radio"/> Língua Portuguesa.</p> <p><input type="radio"/> Matemática e Língua Portuguesa.</p> <p><input type="radio"/> Outra(s) disciplinas.</p> <p><input type="radio"/> Nenhuma disciplina.</p> <p>5.7. Você já evadiu-se de algum curso presencial em outra Instituição sem dar conhecimento à Coordenação do curso? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>5.8. Você já evadiu-se de algum curso presencial do INSTITUTO FEDERAL DO PIAUÍ (IFPI) sem dar conhecimento à Coordenação do Curso? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>5.9. Excluindo você, alguém da sua família já fez algum curso à distância? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>5.10. Você já fez algum Curso à Distância? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>5.11. Você já se evadiu de algum Curso à Distância em outra Instituição sem dar conhecimento à Coordenação do curso? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>5.12. Você já se evadiu de algum Curso à Distância do INSTITUTO FEDERAL DO PIAUÍ (IFPI) sem dar conhecimento à Coordenação do curso? *</p>
<p>5.12. Você já se evadiu de algum Curso à Distância do INSTITUTO FEDERAL DO PIAUÍ (IFPI) sem dar conhecimento à Coordenação do curso? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>5.13. Quem mais lhe incentivou a estudar em um curso a distância? *</p> <p><input type="radio"/> Pais</p> <p><input type="radio"/> Cônjuge / Companheiro(a) / Esposo(a)</p> <p><input type="radio"/> Outros membros da família.</p> <p><input type="radio"/> Colegas/ Amigos</p> <p><input type="radio"/> Ninguém</p> <p>5.14. Qual sua impressão quanto aos cursos à distância? *</p> <p><input type="radio"/> Acredito ser menos exigente que um curso presencial.</p> <p><input type="radio"/> Acredito ser exigente tão quanto um curso presencial.</p> <p><input type="radio"/> Acredito ser mais exigente que um curso presencial.</p> <p><input type="radio"/> Outro: _____</p> <p>5.15. Em sua casa, existe algum cômodo exclusivo para os estudos? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>5.16. Qual o motivo para você ter escolhido este curso? *</p> <p><input type="radio"/> Inserção no mercado de trabalho.</p> <p><input type="radio"/> Influência familiar.</p> <p><input type="radio"/> Valorização profissional.</p> <p><input type="radio"/> Vocação</p> <p><input type="radio"/> Outro: _____</p> <p>5.17. Quais motivos poderiam levar você a se evadir de um curso? OBSERVAÇÃO: Nesta pergunta você pode assinalar mais de uma alternativa. *</p> <p><input type="checkbox"/> Problemas familiares.</p> <p><input type="checkbox"/> Problemas financeiros.</p> <p><input type="checkbox"/> Dificuldades em se relacionar com os colegas de turma.</p> <p><input type="checkbox"/> Dificuldades em se relacionar com o(a) professor(a) / tutor(a) de alguma disciplina.</p> <p><input type="checkbox"/> Dificuldades em lidar com a tecnologia na educação.</p>	<p>5.17. Quais motivos poderiam levar você a se evadir de um curso? OBSERVAÇÃO: Nesta pergunta você pode assinalar mais de uma alternativa. *</p> <p><input type="checkbox"/> Problemas familiares.</p> <p><input type="checkbox"/> Problemas financeiros.</p> <p><input type="checkbox"/> Dificuldades em se relacionar com os colegas de turma.</p> <p><input type="checkbox"/> Dificuldades em se relacionar com o(a) professor(a) / tutor(a) de alguma disciplina.</p> <p><input type="checkbox"/> Dificuldades em lidar com a tecnologia na educação.</p> <p><input type="checkbox"/> Ter dúvida quanto à sua escolha para formação profissional.</p> <p><input type="checkbox"/> Distância da sua casa para o campus ou pólo.</p> <p><input type="checkbox"/> Falta de tempo.</p> <p><input type="checkbox"/> Não saber administrar o tempo para os estudos.</p> <p><input type="checkbox"/> Dificuldades nas disciplinas do curso, devido a pouca base no ensino fundamental e/ou médio.</p> <p><input type="checkbox"/> A Instituição não corresponde às suas expectativas.</p> <p><input type="checkbox"/> Dificuldade de acesso à Internet.</p> <p>6. DIMENSÃO TECNOLÓGICA</p> <p>6.1. Qual o meio de comunicação que você mais utiliza para se manter informado? *</p> <p><input type="radio"/> Livros</p> <p><input type="radio"/> Jornais</p> <p><input type="radio"/> TV</p> <p><input type="radio"/> Rádio</p> <p><input type="radio"/> Internet</p> <p>6.2. Você possui noções básicas de informática? OBSERVAÇÃO: Entendemos por noções básicas em informática: acessar a internet, utilizar o pacote office (word, excel, powerpoint), acessar e-mail, anexar arquivo, fazer download de um arquivo, saber diferenciar hardware de software.</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p> <p>6.2. Antes de ingressar nessa modalidade de ensino, você já possuía e-mail? *</p> <p><input type="radio"/> Sim</p> <p><input type="radio"/> Não</p>

Fonte: Dados da pesquisa (2017).

6.2. Se você já possuía e-mail. Para qual finalidade você o utiliza(va)? *

- Para manter contato diário com colegas de estudo, trabalho e outros.
- Para criar perfil em redes sociais.
- Possuía, mas não fazia uso.
- Não possuía e-mail.

6.3. Você possui computador em casa? *

- Sim
- Não

6.4. Você possui internet em casa? *

- Sim
- Não

6.5. Onde você acessa a Internet, com maior frequência? *

- Em casa.
- No trabalho.
- Lan House.
- Casa de amigos.
- Nunca acesso
- Outro:

6.5. Com qual equipamento você frequentemente acessa a internet? *

- Computador
- Smartphone (celular)
- Tablet
- Não acesso por que não possui nenhum dos equipamentos citados nos itens anteriores.

6.6. O que você sabe sobre a plataforma MOODLE. *

- Não conheço
- Já ouvi falar sobre.
- Já utilizei.
- Outro:

7. DIMENSÃO DE TEMPO

7.1. Quantas horas por semana, aproximadamente, você dedica aos estudos em casa? *

- Uma a três horas por semana.
- Quatro a sete horas por semana.
- Oito a doze horas por semana.
- Mais de doze horas por semana.
- Nenhuma.

7.2. Quantas horas por semana, aproximadamente, você dedica ao trabalho?

- 20 horas.
- De 20 a 40 horas.
- Mais de 40 horas.
- Não trabalho.

7.3. Com relação ao seu tempo, por qual motivo escolheu fazer um curso à distância?

- Não tenho tempo para estudar um curso presencial.
- Economia de tempo com estudo a distância.
- Flexibilidade de horário
- Facilidade de se comunicar para quem é tímido(a).

8. DIMENSÃO DE LOCALIZAÇÃO

8.1. Sua residência fica localizada na zona: *

- Urbana
- Rural

8.2. Com relação a sua localização, por qual motivo você escolheu fazer um curso à distância? *

- Moro distante de instituições que ofertam cursos presenciais.
- Moro distante de instituições que ofertam este curso.
- Economia de tempo com o deslocamento do trabalho para o curso.
- Economia de tempo com o deslocamento de casa para o curso.
- Possibilidade em conciliar com outras atividades do meu dia a dia (família, trabalho, lazer).

Não sou um robô

 HCAPS/CPA

ENVIAR

7.1. Quantas horas por semana, aproximadamente, você dedica aos estudos em casa? *

- Uma a três horas por semana.
- Quatro a sete horas por semana.
- Oito a doze horas por semana.
- Mais de doze horas por semana.
- Nenhuma.

7.2. Quantas horas por semana, aproximadamente, você dedica ao trabalho?

- 20 horas.
- De 20 a 40 horas.
- Mais de 40 horas.
- Não trabalho.

7.3. Com relação ao seu tempo, por qual motivo escolheu fazer um curso à distância?

- Não tenho tempo para estudar um curso presencial.
- Economia de tempo com estudo a distância.
- Flexibilidade de horário
- Facilidade de se comunicar para quem é tímido(a).

8. DIMENSÃO DE LOCALIZAÇÃO

8.1. Sua residência fica localizada na zona: *

- Urbana
- Rural

8.2. Com relação a sua localização, por qual motivo você escolheu fazer um curso à distância? *

- Moro distante de instituições que ofertam cursos presenciais.
- Moro distante de instituições que ofertam este curso.
- Economia de tempo com o deslocamento do trabalho para o curso.
- Economia de tempo com o deslocamento de casa para o curso.
- Possibilidade em conciliar com outras atividades do meu dia a dia (família, trabalho, lazer).

Não sou um robô

 HCAPS/CPA

ENVIAR

Fonte: Dados da pesquisa (2017).