



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

UNIDADE ACADÊMICA DE SERRA TALHADA

BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Um *Dataset* para Avaliação da Qualidade do Ensino Superior a partir de Dados Abertos

Por

Alison Rodrigo Cavalcante de Souza

Serra Talhada,
Dezembro/2021



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
UNIDADE ACADÊMICA DE SERRA TALHADA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

ALISON RODRIGO CAVALCANTE DE SOUZA

Um *Dataset* para Avaliação da Qualidade do Ensino Superior a partir de Dados Abertos

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Unidade Acadêmica de Serra Talhada da Universidade Federal Rural de Pernambuco como requisito parcial à obtenção do grau de Bacharel.

Orientador: Dr. Ellen Polliana Ramos Souza
Coorientador: Ma. Fabiana Pereira Costa

Serra Talhada,
Dezembro/2021

Um *Dataset* para Avaliação da Qualidade do Ensino Superior a partir de Dados Abertos

Alison Rodrigo C. Souza¹, Ellen Polliana R. Souza¹, Fabiana P. Costa¹

¹Universidade Federal Rural de Pernambuco (UFRPE) –
Unidade Acadêmica de Serra Talhada (UAST)
Caixa Postal 063 – 56.900-000 – Serra Talhada – PE – Brazil

{alison.rodrigocavalcante,ellen.ramos}@ufrpe.br, fabiana.pereiracosta@gmail.com

Abstract. *Over the past years Brazil has gone through a phenomenon of expansion for the higher education system. In this context, a large amount of data has been made available, which can be used to support research and assessments. However, those datas are published in a format that is difficult to understand, in addition of being distributed in diferente sources. Thus, this work aims to integrate, structure and make available a dataset with open higher education data from multiple sources. Furthermore, this work presents a visualization model that helps in the decision-making process, with the presentation of higher education quality indicators. The dataset contains data from the Higher Education Census, ENADE, IBGE and e-MEC, from 2009 to 2019.*

Keywords: Quality Indicators, Open Data, Higher Education.

Resumo. *Nos últimos anos o Brasil tem passado pelo fenômeno da expansão do sistema de ensino superior. Nesse contexto, uma grande quantidade de dados tem sido disponibilizada, que pode se utilizar para subsidiar pesquisas e avaliações. No entanto, esses dados são publicados em um formato que dificulta o entendimento, além de serem distribuídos em diferentes fontes. Este trabalho tem como objetivo a integração, estruturação e disponibilização de um dataset com dados abertos do ensino superior, a partir de múltiplas fontes. Além disso, este trabalho apresenta um modelo de visualização que permite auxiliar no processo de tomada de decisão, com a apresentação de indicadores de qualidade do ensino superior. O dataset contém dados do Censo da Educação Superior, ENADE, IBGE e e-MEC, do período de 2009 a 2019.*

Palavras-chave: Indicadores de Qualidade, Dados Abertos, Ensino Superior.

1. Introdução

O Ensino Superior é o nível mais elevado de formação de indivíduos no sistema educacional brasileiro [Costa et al. 2017]. Desde o final dos anos 90, houve uma significativa expansão de cursos superiores no país, liderada principalmente pelo setor privado, seguida de uma expansão em menor escala pelo setor público. O setor privado foi impulsionado pelo financiamento de agentes privados estrangeiros e nacionais, bem como por programas como PROUNI¹ (Programa Universidade para Todos) e pelo fortalecimento do FIES

¹Programa implementado a partir do Decreto Nº 5.493, de 18 de julho de 2005

(Fundo de Financiamento ao Estudante do Ensino Superior). Já o ensino superior público foi apoiado pela implementação de políticas públicas [Moura 2014], como por exemplo, o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI)².

No ensino superior, uma grande quantidade de dados tem sido gerada anualmente. A publicação desses dados ocorre devido a obrigatoriedade de sua disponibilização para acesso público, tais como a Lei de Acesso à Informação (LAI) Lei 12.527/2011 [Brasil 2011]. Dentre os dados educacionais disponibilizados de forma aberta destacam-se aqueles coletados pelo Instituto de Pesquisas Educacionais Anísio Teixeira (INEP)³, responsável pelo Censo da Educação Superior e Exame Nacional de Desempenho dos Estudantes (ENADE)⁴. Ambos fornecem indicadores que permitem avaliar a qualidade do ensino superior brasileiro. Essas bases de dados têm sido usadas por diversos autores, tais como [Bielschowsky 2018], [Souza et al. 2017], [Lacerda and Ferri 2015], para subsidiar estudos acerca do ensino superior brasileiro.

Medir o desempenho de uma instituição de ensino superior é essencial para avaliar a qualidade do ensino fornecido. Isso pode ser realizado por meio de indicadores que forneçam informações capazes de ajudar a encontrar falhas e/ou processos que precisam ser otimizados [Boynard 2013]. Em especial, as bases de dados do INEP são usadas para computo de diversos indicadores de qualidade, tais como Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD), Conceito Preliminar de Curso (CPC) e o Conceito ENADE.

Um dos problemas mais recorrentes que as avaliações de desempenho procuram compreender é a evasão [Kowalski 2020], um fenômeno discutido por diferentes autores na literatura, sendo comumente definida como a saída definitiva do aluno do sistema educacional, a interrupção no ciclo de estudos [Gaios 2005], saída definitiva do aluno do curso de origem sem concluí-lo [Palharini 2010] e/ou saída do aluno da instituição de ensino antes de concluir o curso [Baggi and Lopes 2011].

Essa falta de consenso no entendimento do que é evasão se reflete também nas metodologias de cálculo existentes. É possível calcular a evasão por instituição e Evasão do Sistema, retirando-se do cálculo os ingressantes oriundos de transferências entre Instituições de Ensino Superior (IES), ou entre cursos na mesma IES, ou ainda, considerar a suspensão ou interrupção temporária da matrícula discente como índice para evasão.

Esse desencontro entre os indicadores para o cálculo da evasão, aliado ao crescente volume de dados, são desafios para gestores de instituições de ensino superior monitorarem a qualidade do seu ensino. Desta forma, se torna necessário o uso de soluções com características que se diferenciam dos sistemas de informações tradicionalmente utilizados. Nesse contexto, os Sistemas de Apoio à Decisão (SAD), se tornam importantes, para auxiliar, o processo de descoberta de informação e tomada de decisão.

Assim, este trabalho tem como objetivo a criação e a disponibilização de um *dataset* para avaliação do ensino superior, utilizando dados abertos do ensino superior no período de 2009 a 2019. Para tanto, já foram alcançadas as seguintes contribuições: *i* -

²Programa implementado a partir do Decreto Nº 6.096, de 24 de abril de 2007

³<http://www.inep.gov.br/>

⁴<https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enade>

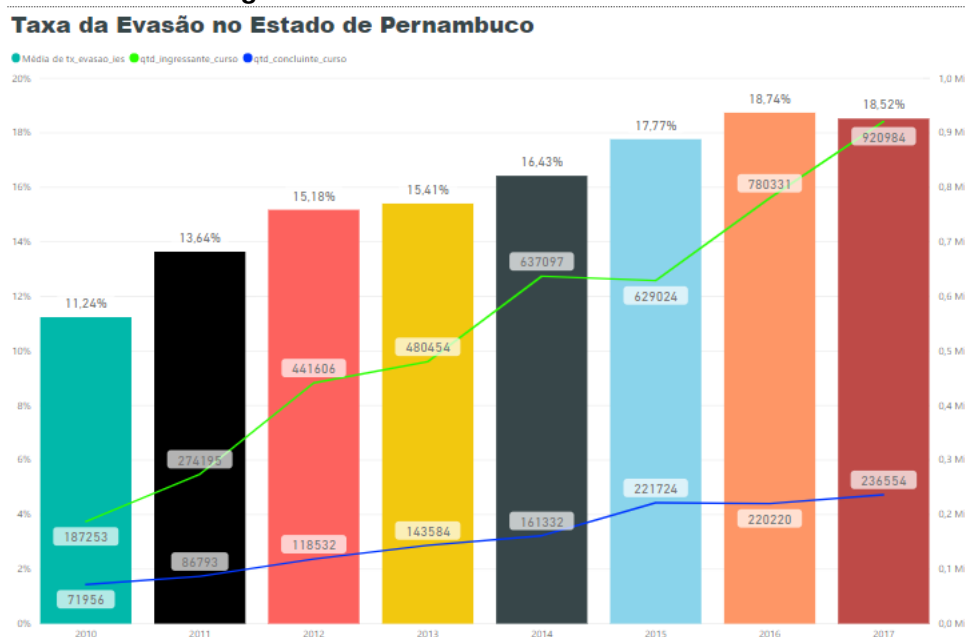
Dataset integrado com indicadores de qualidade e de evasão do ensino superior brasileiro
 ii - Modelo dimensional para análises a partir de várias dimensões e granularidades
 iii - *Online Analytical Processing* (OLAP) dos indicadores de qualidade do ensino superior.
 Este artigo está organizado da seguinte maneira: na Seção 2, são apresentados os trabalhos relacionados. A Seção 3 apresenta o procedimento metodológico. A Seção 4 expõe os resultados e a Seção 5 apresenta a conclusão.

2. Trabalhos Relacionados

O projeto e desenvolvimento de sistemas que apoiam o monitoramento da qualidade do ensino superior tem sido alvo de vários trabalhos, tais como os de [Souza et al. 2017] e [Brito et al. 2014]. Um grande número de sistemas de apoio à decisão tem sido criados, como evasão [Carvalho 2019], indicadores do IDEB [Nascimento et al. 2017] ou inclusão de pessoas com deficiência [Maia et al. 2019]. Incluindo a construção de aplicações Olap como os trabalhos de [Carvalho 2019, Santos 2017, Hoed 2016].

O trabalho de [Carvalho 2019] procurou analisar a evasão nos cursos superiores do estado de Pernambuco. Para isto, os autores apresentaram um *Data Mart*, com dados abertos do Censo da Educação Superior, fornecidos pelo INEP, no período de 2009 a 2017. O processo utilizado foi o proposto por [Kimball and Ross 2011] e adaptado por [Barbieri 2001]. No processo de desenvolvimento de aplicações, a aplicação adotada foi do tipo OLAP, sendo utilizada a ferramenta *Microsoft Power BI*⁵. Para o tratamento dos dados, foi adotada a ferramenta *Pentaho Data Integration*⁶. A partir dos resultados obtidos do trabalho, foi possível visualizar os índices da evasão do ensino superior no estado de Pernambuco, como apresentado na Figura 1, subdividido por categoria administrativa e organização acadêmica, como demonstrado na Figura 2, entre outros.

Figura 1. Evasão no Estado de Pernambuco.



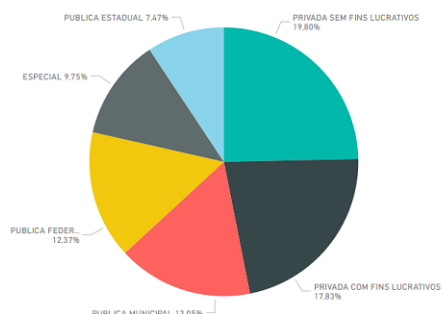
Fonte:[Carvalho 2019]

⁵<https://powerbi.microsoft.com/pt-br/>

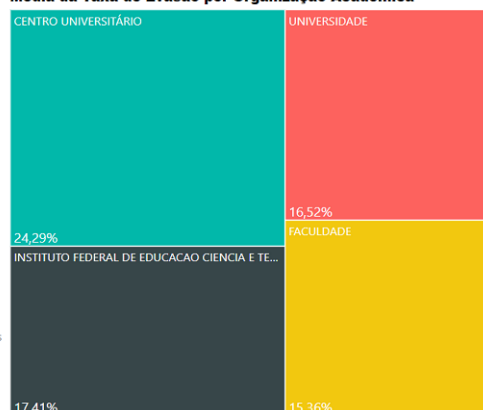
⁶<https://help.hitachivantara.com/Documentation/Pentaho/9.2>

Figura 2. Evasão por Categoria Administrativa e Organização Acadêmica.

Evasão por Categoria Administrativa



Média da Taxa de Evasão por Organização Acadêmica



Fonte:[Carvalho 2019]

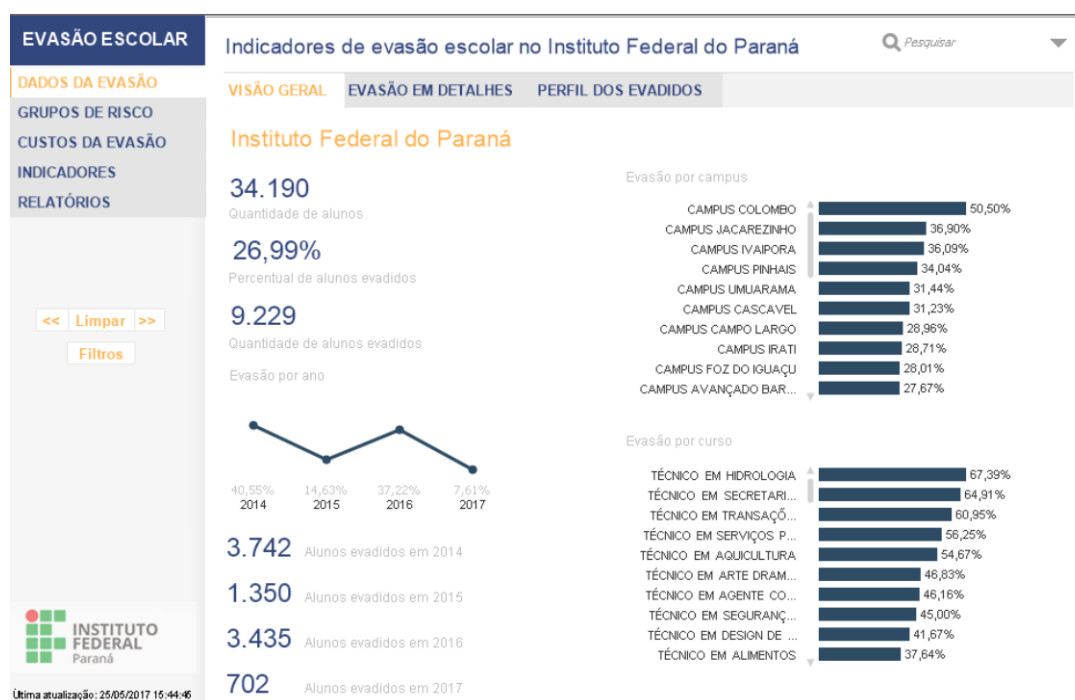
[Santos 2017] propôs um modelo de aplicação de BI como proposta metodológica visando auxiliar o processo de tomada de decisão sobre a evasão escolar em universidades e institutos federais, em especial o Instituto Federal do Paraná (IFPR). Como processo de desenvolvimento da aplicação, o autor aplicou a metodologia de mineração de dados CRISP-DM adaptada ao contexto de BI. Para a implementação da aplicação OLAP o autor adotou a ferramenta de BI QlikView. A aplicação desenvolvida permitiu a visualização geral sobre a evasão na instituição, como apresentado na Figura 3, contendo a evolução da evasão, bem como percentuais por unidade, demonstrado na Figura 4, evadidos correlacionados com o total, entre outros.

Figura 4. Filtros da Evasão no Instituto Federal do Paraná.



Fonte:[Santos 2017]

Figura 3. Evasão no Instituto Federal do Paraná.



Fonte:[Santos 2017]

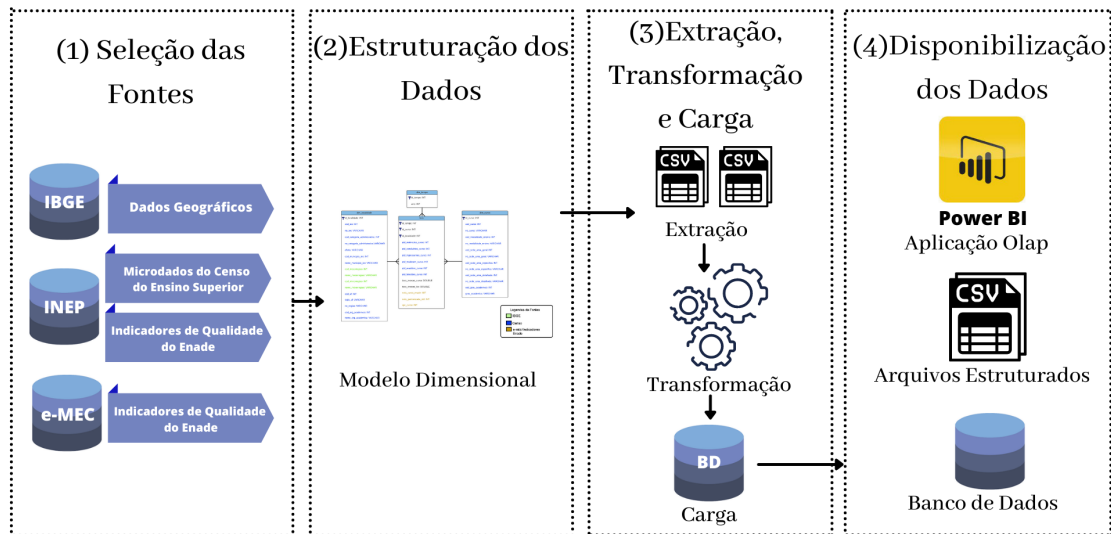
[Hoed 2016] analisou a evasão nos cursos superiores do Brasil, com foco nas causas da evasão nos cursos superiores de Computação na Universidade Federal de Brasília (UNB). O autor, que também utilizou-se da metodologia CRISP-DM, fez uso dos dados do Censo do período de 2010 a 2014. O autor utilizou-se também da metodologia . Na apresentação das taxas de evasão anuais o autor utilizou para geração dos gráficos o SPSS Statistic. O trabalho possibilitou verificar que, de maneira geral, a evasão acontece mais rapidamente no início do curso. Também, encontrou-se evidências que sugerem que nas disciplinas iniciais relacionadas à matemática aumentam a probabilidade do aluno evadir-se.

Para a análise dos indicadores de qualidade do ensino superior, foram encontrados poucos trabalhos que consideram dados em um escopo mais amplo de análise, abrangendo todas as instituições, cursos e áreas do ensino superior brasileiro.

3. Materiais e Método

O método utilizado para a estruturação do *dataset*, modelagem dos dados, construção do *Data Mart* e disponibilização da aplicação OLAP é dividido em quatro etapas principais: seleção das fontes, estruturação dos dados, extração, transformação e carga (ETC), disponibilização dos dados. Baseado no modelo proposto por [Kimball and Ross 2011]. A Figura 5 apresenta a integração dessas etapas.

Figura 5. Metodologia para a Construção do dataset.



A primeira etapa do método teve como objetivo definir as fontes de dados de interesse, com base no levantamento dos indicadores necessários. Essa busca foi realizada por meio de pesquisa bibliográfica. A partir desse levantamento, foram selecionados dados provenientes de diferentes fontes: dados geográficos do Instituto Brasileiro de Geografia e Estatística (IBGE)⁷, indicadores de qualidade e dados do censo (INEP, E-MEC). Além disso, a partir do levantamento, foi definido o cálculo da evasão a ser adotado. Assim, neste estudo, o cálculo da evasão compreendeu tanto o conceito de evasão de curso, como o conceito de evasão de instituição, ambos seguindo as formas de cálculo apresentadas por [Silva Filho and Lobo 2012], demonstradas na Tabela 3. Nas fórmulas, I2 se refere ao número de ingressantes em um determinado ano, M2 corresponde ao número de matrículas neste mesmo ano, C1 é o número de concluintes no ano anterior e M1 o número de matriculados no ano anterior. É válido ressaltar que foi considerado o aluno que muda de curso dentro da instituição, como evadido. Portanto, para calcular a evasão de IES, é necessário retirar dos ingressantes o número de estudantes que mudaram de cursos, mas não de IES (ITC2).

Tabela 1. Tipos de Cálculo para evasão.

Forma de Evasão	Cálculo
Evasão de Curso	$(M2 - I2) / (M1 - C1)$
Evasão de IES	$(M2 - I2 + ITC2) / (M1 - C1)$

A segunda etapa teve como foco a escolha do modelo dimensional e a estruturação dos dados em dimensões que facilitassem as consultas e as integrações realizadas nas etapas posteriores. Foi utilizado o modelo dimensional e o esquema estrela proposto por [Kimball and Ross 2011].

⁷<https://www.ibge.gov.br/pt/inicio.html>

A terceira etapa compreendeu a organização e tratamento dos dados no *dataset*, no qual foi utilizada a ferramenta *Pentaho Data Integration*, integrante da plataforma *Pentaho Business Intelligence*. A seguir, há um detalhamento das subetapas que compõem essa etapa:

Na subetapa de extração, os arquivos são carregados na ferramenta, considerando os delimitadores e a formatação de cada arquivo e são selecionados os dados necessários para os cálculos dos indicadores de interesse que foram levantados.

Já na subetapa de transformação, cada dimensão teve uma transformação específica de acordo com os indicadores nela definidos. Dentre as atividades realizadas, destacam-se: retirar acentos, caixa alta, tratamentos dos campos nulos (Limpeza de Dados). Junção dos indicadores das bases a partir de identificadores (Integração). Separação de IES presencias e EAD, separação de campus e sede (Filtragem dos dados). Redução de valores dos indicadores de qualidade de contínuo para faixa (Condensação dos dados). Realização do cálculo da evasão (Derivação dos dados).

A etapa de carga consistiu no projeto físico do banco de dados e inserção dos dados nele. Para isso, foi utilizado o Sistema de Gerenciamento de Banco de Dados (SGBD) relacional *PostgreSQL*⁸.

A última etapa foi responsável pelo desenvolvimento da aplicação OLAP na ferramenta *Microsoft Power BI* e a disponibilização dos arquivos estruturados, banco de dados e gráficos na web. A escolha das ferramentas adotadas para o desenvolvimento deste trabalho seguiu o uso de ferramentas com versão gratuita, *open source* ou com licenças para estudantes.

4. Resultados

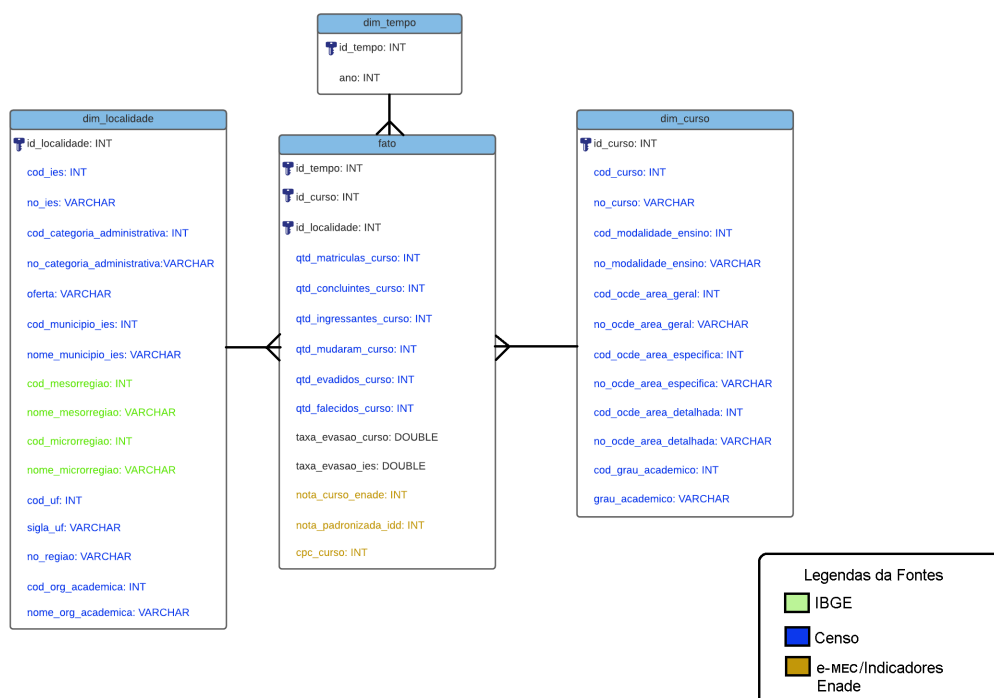
Nas subseções a seguir são apresentados os resultados deste trabalho: A Subseção 4.1 apresenta o modelo, a Subseção 4.2 apresenta o *dataset*, 4.3 apresenta a análise e visualização de dados, 4.4 inconsistências e problemas com os dados abertos e 4.5 comparação com os trabalhos relacionados.

4.1. Modelo

O modelo dimensional seguiu o esquema estrela, caracterizado pela presença de dados redundantes, o que permite um melhor desempenho nas consultas. O modelo está distribuído da seguinte forma: tabelas dimensões de tempo, local, curso e fato. A dimensão tempo possui apenas um atributo o ano. A dimensão local possui dados relacionados a localização da instituição, tais como: nome da região, sigla uf, nome da mesorregião, nome da microrregião, nome do município da IES, nome da IES. A dimensão curso possui dados que caracterizam os cursos tais como: área geral, detalhada e específica, de acordo com a classificação da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), adotada pelo INEP, modalidade de ensino, grau acadêmico e nome do curso. E a dimensão fato que comporta as métricas e os indicadores utilizados nas análises, conforme é apresentado na Figura 6.

⁸<https://www.postgresql.org/>

Figura 6. Modelo dimensional.



Devido à quantidade de atributos e indicadores de diferentes bases do ensino superior, que foram incorporados neste trabalho, uma variedade de análises pode ser feita. As visualizações possíveis são: todas as métricas da dimensão fato; os quantitativos de ingressantes; concluintes; matriculados; evadidos; falecidos; as taxas de evasão de curso e instituição; os indicadores de qualidade IDD; CPC e Conceito ENADE. Todos em separado ou em conjunto, filtrados por qualquer um dos atributos das dimensões curso, local e tempo.

4.2. Dataset

Os dados que compõem o *dataset* foram extraídos do Portal do INEP no período de 2009 a 2019⁹, compreendendo os microdados do Censo da Educação Superior e os dados referentes aos indicadores de qualidade relacionados ao ENADE. Os dados referentes às mesorregiões e microrregiões foram extraídos do servidor do IBGE. Para complementar os indicadores de qualidade, foram usados os dados do Cadastro e-MEC de Instituições e Cursos do Ensino Superior¹⁰, os quais tiveram que ser agregados por estado, pois a consulta dos dados no site ao que concerne ao Brasil era inviável. Portanto, foi necessário realizar consultas separadas e agregar os dados.

⁹<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>

¹⁰<https://emec.mec.gov.br/>

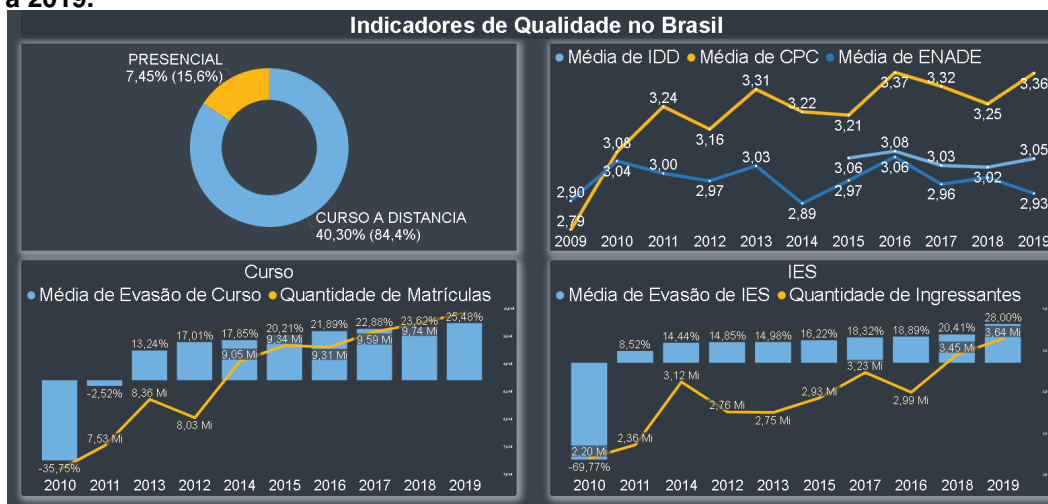
Os microdados do Censo da Educação Superior estão divididos em dimensões, das quais foram utilizadas a de curso, de IES e de aluno. A base de dados do Censo contém um dicionário de dados para cada dimensão, explicando os dados, bem como seus tipos e variedade de representações.

A organização do *dataset* tomou como base o modelo dimensional definido, com os indicadores levantados que servem de filtro para a variedade de análises. A dimensão de curso contém dados referentes às áreas com vários níveis de detalhamento, além de dados sobre grau acadêmico e modalidade de ensino. Já a dimensão de localidade, contém dados que localizam as instituições no espaço geográfico brasileiro e que as classificam de acordo com sua organização acadêmica e categoria administrativa.

4.3. Análise e Visualização de Dados

Foi construída uma aplicação OLAP que permite diversos tipos de visualizações dos indicadores de qualidade. A Figura 7 apresenta o *Dashboard* dos indicadores de qualidade de curso e instituição, no qual é possível observar um crescimento contínuo moderado da evasão escolar ao longo dos anos, com exceção de 2019, pois houve um aumento significativo em relação a 2018, se comparado com os anos anteriores. Com relação aos indicadores de qualidade relacionados ao ENADE, o CPC vem crescendo ao longo do período, em contrapartida, o IDD e o Conceito ENADE estão se mantendo nas suas médias. No que tange à modalidade de ensino, os cursos à distância apresentaram uma média da taxa de evasão, ao longo dos anos do estudo, de 40,3%. Enquanto que a modalidade presencial apresentou 7,45%. É fácil perceber que o valor -35,75% para o ano de 2010 não condiz com a evasão real. No entanto, fica difícil em algumas circunstâncias obter uma precisão no cálculo da evasão quando se trabalha com dados agregados [Hoed 2016]. Portanto, podem surgir alguns problemas para se calcular a evasão usando dados do INEP, como por exemplo, em 2009 foram encontrados cursos com quantidade de matrículas inferiores à quantidade de vagas disponíveis, e em 2010 as vagas foram preenchidas, o que pode explicar a evasão negativa para o respectivo ano.

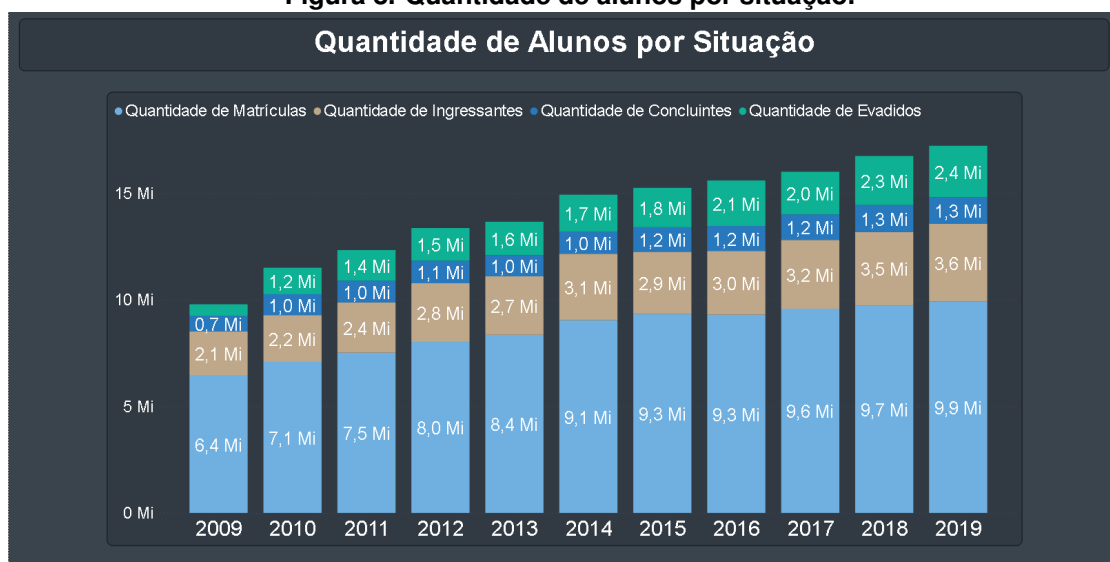
Figura 7. Indicadores de qualidade no ensino superior brasileiro no período de 2009 a 2019.



Na Figura 8 são apresentados os dados quantitativos referentes ao número de ma-

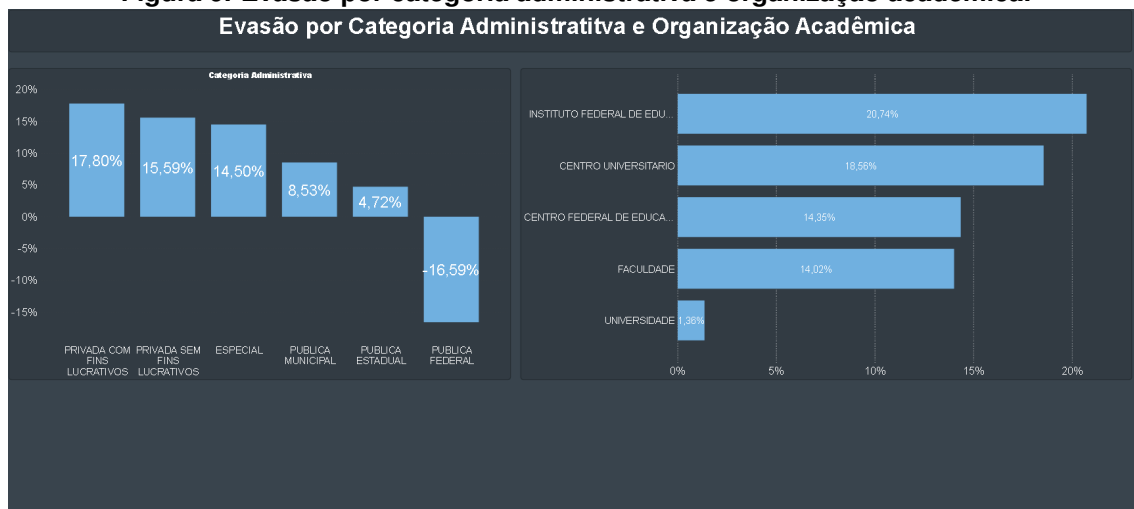
tricolados, ingressantes, concluintes e evadidos. Observa-se uma expressiva quantidade de alunos que se evadem por ano, havendo um maior número em relação aos alunos concluintes. Salientando que, nesse caso, foi considerada a quantidade de evasão tanto de curso como de instituição, ou seja, mesmo que o aluno tenha saído do curso para ingressar em outro, na mesma IES, ele está sendo considerado nesta visualização. O número de matrículas e ingressantes vêm crescendo ao longo dos anos, evidenciando a expansão gradual do ensino superior brasileiro.

Figura 8. Quantidade de alunos por situação.



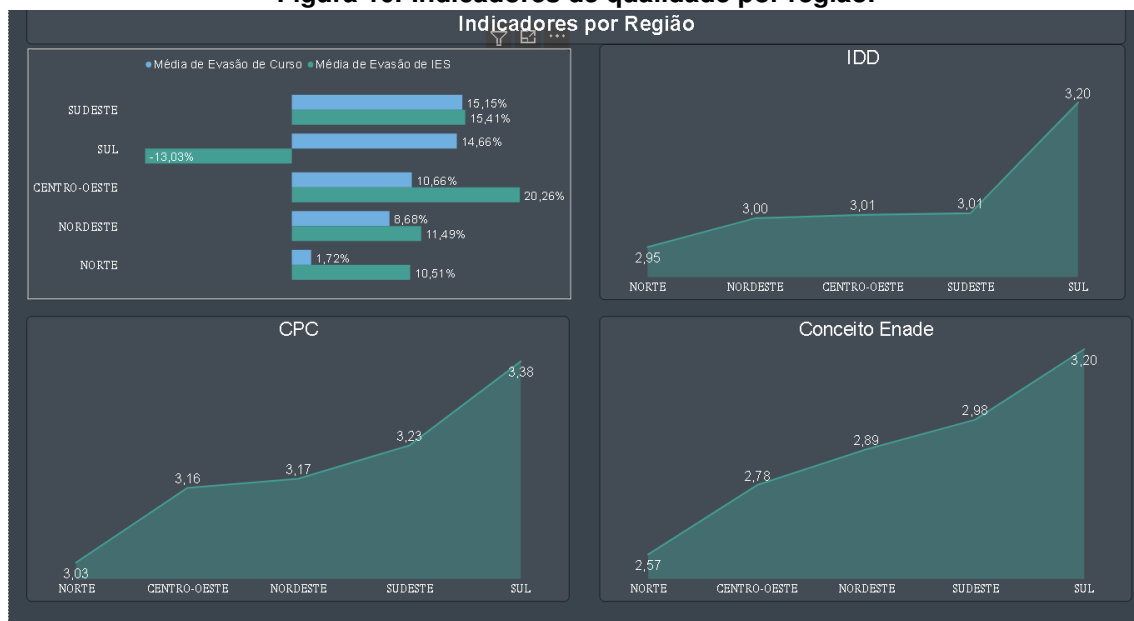
Em relação ao tipo de categoria administrativa, apresentado na Figura 9, a categoria Privada com Fins Lucrativos apresenta o maior percentual de evasão das categorias para o período estudado com 24,37%. No entanto, apesar de as categorias de Instituições Públicas apresentarem menores taxas, vale ressaltar que, quando somadas, apresentam uma elevada taxa de evasão. Isso é um fator preocupante, pois representa um desperdício elevado do dinheiro público investido na formação dos profissionais. No que se refere à organização acadêmica, a de maior taxa média de evasão no período foi o Instituto Federal de Educação e Tecnologia com 20,74%. Centro Universitário, Centro de educação Tecnológica (CEFET), Faculdade e Universidade possuem respectivamente 18,56%, 14,35%, 14,02, 21,36% de taxa média de evasão.

Figura 9. Evasão por categoria administrativa e organização acadêmica.



A Figura 10 apresenta os indicadores de qualidade por região do Brasil. A região Sul apresenta a melhor média dos indicadores de qualidade relacionados com o ENADE, em seguida a região Sudeste. A região Norte apresenta a menor média dos indicadores, já Nordeste e Centro-Oeste permanecem próximas, alternando somente no IDD. Apesar das regiões Sul e Sudeste apresentarem os melhores indicadores, também possuem as maiores taxas de evasão de curso e IES, com exceção da evasão de IES para região Sul, que é negativa.

Figura 10. Indicadores de qualidade por região.



Dado o apresentado, diferentes análises podem ser realizadas acerca da qualidade do ensino superior. Desta forma, uma aplicação OLAP de visualização de dados

pode contribuir no processo de avaliação da qualidade do ensino superior, possibilitando aos gestores uma melhor compreensão dos dados disponibilizados, por oferecerem uma análise em diferentes níveis de visualizações, comparando diferentes indicadores.

4.4. Inconsistências e Problemas com os Dados Abertos

Ao longo do desenvolvimento deste trabalho, foram encontradas algumas inconsistências nos dados abertos do Censo e na base do e-MEC, que é a base de dados oficial e única de informações relativas às IES, servindo como base para o Censo da Educação Superior. Portanto, espera-se que haja conformidade com os dados divulgados no Censo com os disponibilizados para consulta pelo Cadastro e-MEC. As inconsistências encontradas estão exemplificadas na Tabela 2.

Tabela 2. Inconsistências encontradas.

Descrição	Exemplo
Cursos vinculados em IES diferentes nas bases.	Curso A vinculado a IES B no Censo, mas no e-MEC vinculado a IES C.
Cursos presentes em apenas uma base.	Curso A presente no e-MEC, mas não existindo no censo.
Curso presente em um determinado ano em ambas as bases, mas em outro ano existindo somente em uma das bases e mudando de IES para outra.	Curso A vinculado a IES B no Censo e no e-MEC no ano X. No ano X+1 vinculado a IES B no e-MEC, mas no Censo passa a ser vinculado em IES C.
Mudança no agregamento dos cursos EAD.	No ano de 2009 os cursos EAD eram agregados em um só arquivo junto com os presenciais.
Acréscimo de uma categoria administrativa.	Em 2010 foi acrescentado a categoria administrativa especial.
Acréscimo de uma categoria na organização acadêmica.	Em 2010 foi acrescentado à organização acadêmica: Centro Federal de Educação Tecnológica.
Retirada de colunas referentes a nomes de alguns atributos.	No ano de 2017 os nomes das áreas OCDE foram retiradas, permanecendo somente os códigos referentes.
Substituição dos códigos referentes a áreas OCDE.	No ano de 2018 a classificação OCDE foi substituída pelo código cine_rótulo.

As inconsistências encontradas nas bases disponibilizadas podem impactar na confiabilidade dos dados disponibilizados pelo INEP. No geral, entre os problemas já mencionados, destacam-se a falta de completude, qualidade da informação e valores com padrões diferentes.

4.5. Comparação com os Trabalhos Relacionados

Como forma de validar os dados do *dataset* e a aplicação disponibilizada, foram comparados os resultados do estudo com os trabalhos relacionados. Vale ressaltar que, devido

a complexidade do estudo da evasão e sua variedade de definições e diferentes tipos de cálculos, os indicadores podem sofrer variações [Lima and Zargo 2018].

[Hoed 2016] utilizou os dados do INEP no período de 2010 a 2014. Por isso, foram comparados os valores correspondentes a estes anos. Em relação a evasão por área geral, [Hoed 2016] separou a análise em instituições públicas e privadas. Tanto nas instituições públicas quanto nas privadas, a área geral de serviços classificada de acordo com a classificação OCDE, foi a que possuiu a maior taxa de evasão, para o estudo, com 21,50% nas privadas e 16,07% nas públicas. Neste trabalho, a área de serviços também possui a maior taxa anual média de evasão, com 14,25% nas privadas e 13,78% nas públicas.

Já [Santos 2017] analisou especificamente o caso do Instituto Federal do Paraná no período de 2014 a 2017, considerando todos os *campi*, apresentando uma média de evasão de 26,99% para o período. Observando separadamente, a evasão vem diminuindo, saindo de 40,55% em 2014 para 7,61% em 2017. Neste trabalho, a média de evasão para o período no IFPR foi de 23,49%, e ao contrário do trabalho citado, a evasão vem aumentando, saindo de 22% em 2014 para 27,33% em 2017.

[Carvalho 2019] considerou o período de 2009 a 2017, analisando o estado de Pernambuco e apresentou um crescimento contínuo da evasão para o período, atingindo 18,58% em 2017. Em relação a categoria administrativa, as categorias privadas apresentaram a maior média de evasão. Entretanto, foi ressaltado que, as IES públicas quando somadas, apresentam elevadas taxas de evasão. Analisando as modalidade de ensino, a modalidade a distância apresenta a maior taxa de evasão com 21,98% e a presencial com 13,50%. Este trabalho, também apresenta um crescimento da evasão para o estado de Pernambuco, atingindo 18,65%. Em relação a categoria administrativa, também foi verificado que as privadas apresentaram a maior média de evasão, mas as públicas, como relatado, se somadas apresentam altas taxas de evasão.

5. Conclusão

Os dados abertos do ensino superior possibilitam uma variedade de análises acerca da qualidade do ensino, como por exemplo, os indicadores aqui apresentados. Ao longo do desenvolvimento deste trabalho foi possível estruturar um *dataset*, com os dados provenientes de bases de dados abertas do Ensino Superior, visando avaliar os indicadores de qualidade do ensino superior brasileiro. Os dados foram visualizados através de uma aplicação OLAP, permitindo avaliar os indicadores de qualidade do ensino superior sob diferentes níveis de abrangência, além disso foi possível constatar e ratificar o que a literatura menciona sobre os problemas enfrentados ao se trabalhar com dados abertos.

Portanto, diferentes visualizações podem ser realizadas acerca da qualidade do ensino superior. Com isso, uma aplicação OLAP pode contribuir no processo de análise do ensino superior, auxiliando os gestores a possuírem uma melhor compreensão dos dados disponibilizados pelas bases de dados, por permitir diferentes tipos de visualizações e comparações com diferentes tipos de dados.

No entanto, a complexidade das variadas análises e possíveis indicadores relacionados com a qualidade ou desempenho do ensino superior brasileiro, não foram esgotados no presente estudo. Desta forma, como trabalhos futuros, pretende-se fazer uso de

mineração de dados, com o objetivo de identificar padrões e relações no *dataset* construído ao fim do projeto.

Referências

- Baggi, C. A. D. S. and Lopes, D. A. (2011). Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 16:355–374.
- Barbieri, C. (2001). *BI-business intelligence: modelagem e tecnologia*. Axcel Books.
- Bielschowsky, C. E. (2018). Análise dos resultados do exame nacional de desempenho de estudantes (enade) para educação a distância do ciclo 2015 a 2017. *EaD em Foco*, 8(1).
- Boynard, K. M. S. (2013). Indicadores de gestão em conflito com indicadores de qualidade?: lições econômicas para a gestão universitária.
- Brasil (2011). Lei nº 12.527, de 18 de novembro de 2011.
- Brito, D. M. D., de Almeida Júnior, I. A., Queiroga, E. V., and Do Rêgo, T. G. (2014). Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 25, page 882.
- Carvalho, J. A. D. (2019). Sistema de apoio à decisão para análise da evasão no ensino superior. *Serra Talhada, PE: Universidade Federal Rural de Pernambuco*.
- Costa, F. J., Costa, P. R. S., de. Jr. Moura, P. J., and (2017). *DIPLOMAÇÃO, EVASÃO E RETENÇÃO: MODELO LONGITUDINAL DE ANÁLISE PARA O ENSINO SUPERIOR*, volume 73.
- Gaioso, N. P. d. L. (2005). O fenômeno da evasão escolar na educação superior no brasil. *Brasília, DF: Universidade Católica de Brasília*, page 20.
- Hoed, R. M. (2016). Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de computação. *Brasília, DF: Universidade de Brasília*.
- Kimball, R. and Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- Kowalski, A. R. (2020). Propostas para minimizar a evasão em educação a distância.
- Lacerda, L. L. V. d. and Ferri, C. (2015). Relações entre indicadores de qualidade de ensino e desempenho de estudantes dos cursos de pedagogia do brasil no exame nacional de desempenho dos estudantes. *Revista Brasileira de Estudos Pedagógicos*, 96:129–145.
- Lima, F. S. D. and Zargo, N. (2018). Desafios conceituais e tendências da evasão no ensino superior: a realidade de uma universidade comunitária. *Revista Internacional de Educação Superior*, 4(2):366–386.
- Maia, D., Souza, E. P. R., Oliveira, M. I. S., and (2019). Uma análise da inclusão escolar de pessoas com deficiência a partir de dados governamentais abertos. *Simpósio Brasileiro de Banco de Dados*, pages 317–324.

- Moura, D. M. d. (2014). Políticas públicas educacionais prouni e fies: democratização do acesso ao ensino superior. *Seminário Internacional Demandas Sociais e Políticas Públicas na Sociedade Contemporânea*, (11).
- Nascimento, L. Y. D. B., da Silva. Pereira, V., Souza, E. P. R., and (2017). Data mart para análise comparativa de dados do ideb em municípios da microrregião do pajeú em pernambuco. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 704.
- Palharini, F. d. A. (2010). Evasão, exclusão e gestão acadêmica na uff: passado, presente e futuro. *Niterói: ICHF*.
- Santos, J. S. D. (2017). Business intelligence: uma proposta metodológica para análise da evasão escolar em instituições federais de ensino.
- Silva Filho, R. L. L. and Lobo, M. B. d. C. M. (2012). Esclarecimentos metodológicos sobre os cálculos de evasão. *Instituto Lobo*.
- Souza, H. V. L. D., Neiva, D., Cavalcanti, R. P., Rodrigues, R., Gomes, A. S., and Adeodato, P. (2017). Uma análise preditiva de desempenho dos alunos dos cursos no enade com base no perfil socioeconômico e de desempenho no enem. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 684.

AGRADECIMENTOS

Agradeço primeiramente a Deus, que até aqui me ajudou, aos meus pais, que deram todo o apoio e suporte para me proporcionar o melhor, e sempre estiveram ao meu lado.

Aos meus professores que tanto contribuíram com todos os ensinamentos durante essa jornada, em especial a minha professora orientadora Ellen Polliana Ramos Souza, por toda paciência, atenção e orientação.

Por fim, meus amigos e colegas de curso que contribuíram diretamente ou indiretamente para que fosse possível essa conquista.