

**UNIVERSIDADE DO ESTADO DE MATO GROSSO  
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO  
MESTRADO EM EDUCAÇÃO**

**MARCOS PAULO DE MESQUITA**

**O Perfil dos Estudantes Retidos na Universidade do Estado de Mato Grosso:  
Uma Análise Baseada em Clusterização**

**CÁCERES, MT  
2022**

**MARCOS PAULO DE MESQUITA**

**O Perfil dos Estudantes Retidos na Universidade do Estado de Mato Grosso:  
Uma Análise Baseada em Clusterização**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Educação da Universidade do Estado de Mato Grosso, como requisito parcial para obtenção do título de Mestre em Educação.

Orientador: Dr. Fernando Cezar Vieira Malange

**CÁCERES, MT  
2022**

Luiz Kenji Umeno Alencar CRB 1/2037

M578o	<p>MESQUITA, Marcos Paulo. O Perfil dos Estudantes Retidos na Universidade do Estado de Mato Grosso: Uma Análise Baseada em Clusterização / Marcos Paulo Mesquita - Cáceres, 2022. 63 f.; 30 cm. (ilustrações) Il. color. (não)</p> <p>Trabalho de Conclusão de Curso (Dissertação/Mestrado) - Curso de Pós-graduação Stricto Sensu (Mestrado Acadêmico) Educação, Faculdade de Educação e Linguagem, Câmpus de Cáceres, Universidade do Estado de Mato Grosso, 2022. Orientador: Fernando Cezar Vieira Malange</p> <p>1. Retenção no Ensino Superior. 2. Clusterização. 3. Dados Educacionais. I. Marcos Paulo Mesquita. II. O Perfil dos Estudantes Retidos na Universidade do Estado de Mato Grosso: Uma Análise Baseada em Clusterização: .</p> <p>CDU 378-051(817.2)</p>
-------	---

**Marcos Paulo de Mesquita**

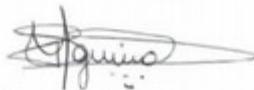
**O Perfil dos Estudantes Retidos na Universidade do Estado de Mato Grosso:  
Uma Análise Baseada em Clusterização**

Dissertação de Mestrado aprovada no Programa de Pós-Graduação em Educação da Universidade do Estado de Mato Grosso, para obtenção do título de Mestre em Educação.

**BANCA EXAMINADORA**



**Dr. Fernando Cezar Vieira Malange**  
(Orientador – PPGEdu/UNEMAT)



**Dr.ª Patrícia Simone Nogueira**  
(Membro Externo – UFMT)



**Dr.ª Elizeth Gonzaga dos Santos Lima**  
(Membro Interno – PPGEdu/UNEMAT)

**APROVADA EM: 30/05/2022.**

*Dedico essa dissertação à minha família, de forma muito especial à minha esposa Marcela, aos meus filhos Melissa e Bento, aos meus familiares que estão em Minas Gerais. E em nome do meu amigo Nivaldi, dedico também esse trabalho à família de amigos que Deus me presenteou aqui em Cáceres.*

*Jamais chegaria sozinho aqui, impossível elencar os agradecimentos sem incorrer à ingratidão de esquecer de alguém. Então, em nome do meu amigo professor orientador, Fernando Cezar Vieira Malange, externo meus mais sinceros agradecimentos a você que agora lê esse texto e entende o que ele realmente significa para mim. Obrigado.*

*“Preso a canções  
Entregue a paixões  
Que nunca tiveram fim  
Vou me encontrar  
Longe do meu lugar  
Eu, caçador de mim.”*

*Milton Nascimento*

## RESUMO

O presente trabalho traz uma abordagem baseada em clusterização de dados educacionais como proposta alternativa para compreender o fenômeno da retenção nos cursos de graduação da Universidade do Estado de Mato Grosso (UNEMAT) analisando os diferentes perfis dos estudantes retidos na instituição. O problema de retenção afeta todas as instituições brasileiras de ensino superior e, na UNEMAT, também se apresenta como uma questão latente. Mitigar a retenção pode significar diminuir os índices de evasão e, conseqüentemente, aumentar as taxas de diplomação. Para tal, o desenvolvimento de políticas institucionais e intervenções pedagógicas mais pontuais dependem do conhecimento do perfil do estudante retido. Este trabalho se configura como um Estudo de Caso, nossos sujeitos foram os estudantes retidos na UNEMAT no ano de 2019 dos cursos regulares de graduação. Constituímos um banco de dados de 21.786 registros a partir do censo da educação superior de 2020, que posteriormente foram enriquecidos com informações do sistema acadêmico da instituição. Neste trabalho consideramos retido o estudante que não obtivera diploma ainda que passado o seu tempo mínimo regulamentar de conclusão de curso. Dessa definição foi possível identificar no nosso banco de dados 2.169 acadêmicos nessa condição. Deste ponto, 3 (três) perfis de retenção foram obtidos por meio do algoritmo de clusterização *K-Modos* e os resultados demonstram no cluster 1 uma considerável incidência de estudantes retidos nos cursos de bacharelados da grande área de Negócios, Administração e Direito, no cluster 2, estudantes das licenciaturas da grande área de Educação e majoritariamente do sexo feminino e no cluster 3 predomina estudantes de turno integral e do sexo masculino, todos em bacharelados. Por fim, foi possível identificar um comportamento de persistência dos estudantes retidos da UNEMAT já que eles, em grande maioria, têm mais de 80% de carga horária cumprida e não apresentam problemas severos de notas.

**Palavras-chave:** retenção no ensino superior; clusterização; dados educacionais

## ABSTRACT

*This master thesis presents an educational data clustering approach as an alternative proposal for attrition comprehension at the undergraduate courses of Universidade do Estado de Mato Grosso (UNEMAT) by analyses of different retained student profiles. The attrition problem affects all Brazilian higher education institutions and, at UNEMAT, it also presents itself as a latent issue. Mitigating retention can mean decreasing dropout rates and, consequently, increasing graduation rates. We think institutional politics and more specific pedagogic interventions depend on having the retained student profile. This research is a case study, we study the retained students from year 2019. A database with 21,786 records was build and afterwards was enriched by data from institutional academic system. In this research, we consider the retained student who has not obtained a diploma even if the minimum regulatory time for completing the course has passed. From this definition, it was possible to identify 2,169 academics in this condition in our database.3 (three) retention profiles were obtained by the K-Modes clustering algorithm and the results demonstrate in the cluster 1 a considerable incidence of students retained in bachelor courses of the area of Business, Administration and Law, in the cluster 2, licensure courses's students of area of Education , mostly female and, in the cluster 3, full-time male students. Also, in the cluster 3, all of students are bachelor courses. Finally, it was possible to identify a behavior of persistence of students retained from UNEMAT since they, for the most part, have more than 80% of the course load fulfilled and do not have severe problems with grades.*

**Keywords:** *high education attrition; clustering; educational data.*

## **LISTA DE SIGLAS E ABREVIATURAS**

COVEST – Comissão Permanente de Concursos e Vestibulares

CESUP – Censo da Educação Superior

CRA – Coeficiente de Rendimento Acadêmico

CINE - Classificação Internacional Normalizada da Educação Adaptada para Cursos de Graduação e Sequenciais de Formação Específica (Cine Brasil)

ENEM - Exame Nacional do Ensino Médio

FACET - Faculdade de Ciências Exatas e Tecnológicas

FIES - Programa de Financiamento Estudantil

KDD – Knowledge Discover in Database

GPAPES – Grupo de Pesquisa em Acesso e Permanência na Educação Superior

IF – Instituto Federal de Ciência e Tecnologia

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

MEC - Ministério da Educação

MED - Mineração de Dados Educacionais

PPC - Projeto Político Pedagógico de Curso

PPGEdu - Programa de Pós-graduação em Educação

PRAE - Pró-Reitoria de Assuntos Estudantis

PROEG - Pró-Reitoria de Ensino de Graduação

PROUNI - Programa Universidade para Todos

REUNI - Reestruturação e Expansão das Universidades Federais

SAGU - Sistema Aberto de Gestão Unificada

SIGAA - Sistema Integrado de Gestão de Atividades Acadêmicas

SISU - Sistema de Seleção Unificada

UNEMAT - Universidade do Estado de Mato Grosso

## LISTA DE TABELAS

<b>Tabela 1</b> - Situação dos estudantes da UNEMAT em 2019.....	43
<b>Tabela 2</b> - Evolução das vagas, ingressos, matrículas e concluintes no ensino superior brasileiro.....	44
<b>Tabela 3</b> - Evolução das vagas, ingressos, matrículas e concluintes na UNEMAT.....	45
<b>Tabela 4</b> - Distribuição dos Evadidos em função do ano de ingresso.....	47
<b>Tabela 5</b> - Distribuição dos Retidos em função do ano de ingresso.....	47

## LISTA DE QUADROS

<b>Quadro 1</b> - Descrição das Variáveis.....	24
<b>Quadro 2</b> - Primeiro Levantamento de Produções na área.....	29
<b>Quadro 3</b> - Sumarização dos trabalhos correlacionados.....	30
<b>Quadro 4</b> - Centróides dos Clusters.....	49

## LISTA DE FIGURAS

<b>Figura 1</b> - O Algoritmo <i>K-Modes</i> .....	27
<b>Figura 2</b> - Termos recorrentes em trabalhos de retenção.....	32
<b>Figura 3</b> - O Modelo de Integração de Tinto.....	41
<b>Figura 4</b> - Relação Concluintes/Matrículas.....	45
<b>Figura 5</b> - Projeção Relação Concluintes/Matrículas.....	46
<b>Figura 6</b> - Determinando o número de Clusters.....	48
<b>Figura 7</b> - Distribuição de Frequência das Variáveis em cada Cluster .....	51

## SUMÁRIO

1	INTRODUÇÃO.....	17
1.1	Objetivos.....	20
2	METODOLOGIA.....	22
2.1	Clusterização.....	25
2.2	O Algoritmo <i>K-Modes</i> .....	26
3	REVISÃO DE LITERATURA.....	29
4	RETENÇÃO: Conceitos e Modelos Teóricos de Estudo.....	38
4.1	O Modelo de Retenção de Vincent Tinto.....	40
5	RETENÇÃO NA UNEMAT.....	43
5.1	Clusterização dos Estudantes Retidos.....	48
6	CONSIDERAÇÕES FINAIS.....	56
	REFERÊNCIAS.....	58

# 1 INTRODUÇÃO

Essa pesquisa é o prelúdio de uma jornada do eu pesquisador. Um caminho rico a ser percorrido, um caminho de diálogos, de trocas e constantes reflexões. Pela minha formação tecnicista em Ciência da Computação, pautada em princípios positivas, sinto-me, não raramente, um estrangeiro nesta trilha, nela não existem atalhos, e eles também não seriam bem-vindos. Por sorte comungo, de certa forma, da língua desse terreno de pesquisa, professo a docência nesta instituição e por observar e experimentar algumas questões no dia a dia do meu labor é que essa pesquisa começa a ter seus primeiros delineamentos: a retenção na Universidade do Estado de Mato Grosso (UNEMAT).

Apresentando-me, sou um pesquisador que acredita que a tecnologia da informação pode ajudar os homens e as mulheres a se tornarem mais humanos na medida que os computadores oferecem recursos ímpares e potentes para muitas das faculdades do ser, em especial as de se comunicar, de se expressar, de interagir. A tecnologia não se desvincula, de modo autônomo, da cultura, nem tão pouco da sociedade; a tecnologia é parte do ambiente material humano, do qual deste não pode ser separado, e, por vezes, aparece retroalimentando e reconstituindo aquele, como discutido em Lévy (1999) e Morin (2010) Reservamos, por hora, qualquer discussão acerca da ética deste contexto, não por ignorar ou minimizá-la, mas pela restrição do escopo da pesquisa.

Também, a tecnologia como ferramenta de pesquisa, de novas possibilidades e novos olhares sobre questões recorrentes, vez ou outra abordada. E isso que o presente trabalho faz; traz uma abordagem baseada em clusterização<sup>1</sup> de dados educacionais<sup>2</sup> como proposta alternativa para estudar o fenômeno de retenção na UNEMAT. E nesse ponto realçamos nossa questão problema que permeia essa dissertação: Por meio da Clusterização de dados educacionais é possível categorizar os diferentes perfis de estudantes retidos na UNEMAT?

---

1 Clusterização ou agrupamento, é uma técnica de aprendizado de máquina que consiste em criar grupos de elementos semelhantes entre si.

2 Dados educacionais designa qualquer dado ou informação de estudantes gerados por sistemas informatizados de ensino, pesquisa, extensão, gestão de ensino, governo e ambientes virtuais de aprendizado.

O ensino superior brasileiro experimentou tanto uma franca expansão na oferta de vagas na década de 1990 e no início dos anos 2000, bem como um fomento de ocupação de vagas nas universidades particulares. Como propulsores deste incremento podemos citar o Programa de Financiamento Estudantil (FIES), o Programa Universidade para Todos (PROUNI), a Reestruturação e Expansão das Universidades Federais (REUNI), a criação dos Institutos Federais de Ciência e Tecnologia (IFET). As políticas de reserva de vagas vão promover também maior acesso ao ensino superior àqueles que historicamente estiveram marginalizados desse nível, configurando-o, pelo menos em tese, como um ambiente plural e democrático.

A despeito desta expansão, as políticas de permanência na universidade brasileira não são implementadas na mesma proporção da oferta de novas vagas. Muitos estudantes por razões diversas acabam por não concluir seus estudos, como revelado por Nunes e Pereira (2019), entre os anos de 2006 e 2016, houve um salto importante no número de não-concluintes (69,06%), nas condições de estudantes retidos ou evadidos no ensino superior do Brasil.

Alinham-se ao cenário nacional os números da UNEMAT de vagas ofertadas, ingressos e evasão e, neste último, podendo em alguns cursos atingir o alarmante índice de 50% (HOFFMANN; BITENCOURT, 2019; NODARI; LIMA; MACIEL, 2018).

O Censo da Educação Superior (CENSUP) de 2020 (INEP, 2020) revelou que ano de 2019, 21.786 estudantes possuíam vínculo com algum dos cursos presenciais de graduação da UNEMAT. Destes, 2.120 diplomaram, 2.501 se desvincularam da instituição e 2.455 já haviam ultrapassado o tempo regulamentar mínimo de integralização de seus cursos. O censo apontou também que o tempo médio de duração dos cursos da instituição fora 4,5 anos, ao passo que os estudantes diplomados estudaram em média 5,57 anos na instituição. Uma parcela de 810 dos estudantes que se desvincularam em 2019, estiveram na universidade por pelo menos 4 anos.

Desses dados, vemos que a Universidade do Estado de Mato Grosso tem questões de retenção latentes e que seria razoável supor que as taxas de diplomação da universidade seriam maiores se as retenções pudessem ser

convertidas em diplomação, ou ainda, se as evasões pudessem ser convertidas em permanência e depois futura conclusão de curso.

Sabemos que não é trivial explicar o porquê dessas retenções ou discutir modelos preditivos de retenção, mas conhecer o perfil daqueles estudantes que se encontram nessa situação é um passo importante para melhor direcionar esforços de gestão, políticas pedagógicas e estratégias de mitigação de evasão e/ou retenção.

O Grupo de Pesquisa em Acesso e Permanência na Educação Superior (GPAPES) vinculado ao PPGEdU/UNEMAT, no qual estamos inseridos, há muito tem apresentado resultados importantes acerca destas temáticas na UNEMAT. Nosso trabalho aparece com uma proposta inédita de análise multivariada, baseada em clusterização de dados educacionais.

A revisão bibliográfica revelou que as produções de pós-graduação feitas nos últimos 6 anos no Brasil acerca de retenção, na maioria das vezes, estão também atreladas a estudos de evasão, que por sua vez aparece como possível causa daquela. Nota-se também que o caráter dos trabalhos é de investigação de causas e ou fatores intrínsecos de retenção nas instituições, com escopo que se fecham em cursos ou em campus. A revisão revelou, ainda que raras, pesquisas que propõe métodos preditivos de evasão.

Nossa pesquisa aborda o estudo de retenção em uma universidade multicampi, que oferece diferentes cursos nas mais diversas áreas do conhecimento, atendendo dezenas de cidades do estado de Mato Grosso. Ao não nos preocuparmos com o caráter preditivo dos dados, ou explicativo do fenômeno, nosso foco é analisar os perfis dos estudantes que atualmente se encontram retidos na instituição por meio de clusterização de dados.

Essa técnica tem por base, analisar um conjunto de dados, que no nosso caso são os registros de estudantes, e agrupá-los em função de suas similaridades. Entre as vantagens, como apontado por Betarelli Junior (2016) estão a possibilidade de avaliar a dimensionalidade, identificar *outliers*<sup>3</sup> e sugerir hipóteses acerca das

---

3 Os *outliers* são dados que se diferenciam drasticamente de todos os outros. Em outras palavras, um outlier é um valor que foge da normalidade e que pode (e provavelmente irá) causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise.

estruturas de relações. A clusterização também é uma importante técnica exploratória de dados.

Mas tão somente a constituição dos clusters não nos habilita avançar nas questões de retenção. Historicamente, o termo tem sido constituído, e a análise dos clusters deve ser feita à luz dessas teorias que tentam dar conta das multifacetadas do fenômeno.

Diferentes abordagens e perspectivas têm sido utilizadas nos modelos teóricos de retenção de estudantes nas últimas décadas com foco nas altas taxas de evasão e retenção de estudante na Educação Superior. Muitas delas parecem ser pouco conhecidas dos educadores, administradores e formuladores de políticas das instituições de ensino em função das suas complexidades. O modelo teórico de Tinto (1975), amplamente estudado e tomado como base para diversas outras propostas, argui que o processo de retenção se dá por meio das interações dos estudantes com os sistemas que ele vai chamar de sistemas institucionais e acadêmicos tendo como base os compromissos e objetivos dos estudantes.

Disso emergiu nossa hipótese de pesquisa de que uma base de dados, que contivesse diferentes variáveis abstraído as dimensões do modelo, poderia ser usada para categorizar os estudantes retidos na instituição.

As técnicas de clusterização auxiliam a identificação de grupos de estudantes retidos que comungam de um perfil, que se assemelham. Dessa forma, no tratamento de retenção, políticas mais bem formuladas, intervenções pontuais e eficientes, e planejamento estratégico tornam-se possíveis para a instituição.

## **1.1 Objetivos**

### ***Objetivo Geral:***

Compreender o fenômeno da retenção nos cursos de graduação da Universidade do Estado de Mato Grosso analisando os diferentes perfis dos estudantes retidos na instituição.

**Objetivos Específicos:**

Consolidar diferentes fontes de informações institucionais para a criação da base de dados de estudantes retidos;

Apresentar os diferentes perfis de estudantes que permanecem retidos na Universidade do Estado de Mato Grosso por meio de clusterização de dados.

Identificar as características que melhor descrevem o processo de retenção dos estudantes.

Dito isso, o texto desta dissertação segue da seguinte forma: a seção 2 descreve os aspectos metodológicos da pesquisa. Ali detalhamos a abordagem do trabalho, apresentamos as etapas seguidas na clusterização. Também definimos e explicamos clusterização e os detalhes do algoritmo *K-Modes* são explorados. A criação do banco de dados também é explicado na seção 2. A seção 3 apresenta nossa revisão de literatura. Apresentamos os trabalhos correlacionados a nossa pesquisa, descrevemos como se configuram hoje as pesquisas sobre o tema retenção por meio de um balanço de produção. A seção 4 discute a teoria da retenção. Investigamos o conceito de retenção, como ele surge e como ele tem se desenvolvido no Brasil. Essa seção também traz o conceito de Modelo Teórico de Retenção e explora o Modelo Teórico de Vincent Tinto. Na seção 5 trazemos os resultados obtidos pela clusterização. Apresentamos alguns indicadores da instituição, como taxas de diplomação, retenção e evasão. Aqui os clusters são apresentados e discutidos. Finalizamos então nossa dissertação destacando os objetivos que alcançamos com nossa pesquisa e com algumas propostas de trabalhos futuros.

## 2 METODOLOGIA

Esse trabalho se constitui, a rigor, de um estudo de caso. Triviños (1987) define estudo de caso como uma categoria de pesquisa cujo objeto é uma unidade que se analisa profundamente. Embora não seja uma classe de pesquisa típica do modelo positivista, mas por suas características que flutuam entre métodos quantitativos e qualitativos, tem sido amplamente explorado nas pesquisas educacionais.

Afastando o debate sobre o antagonismo epistemológico das tendências quantitativa e qualitativa, Gentil (2011) ao discutir Gamboa (2007) destaca o caráter complementar dessas abordagens em prol do rigor e da qualidade da pesquisa que ao levar em conta aqueles dois aspectos, traz uma compreensão e poder preditivo mais considerável. Pereira (2004) também destaca que a pesquisa qualitativa não está isenta de quantificação, nem a pesquisa quantitativa prescinde de raciocínio lógico e que a denominação qualitativa ou quantitativa não delimita para uma e outra, objetos qualitativos e quantitativos.

Por adoção dessa categoria, e pelas especificidades que emergem das pesquisas em retenção, o trabalho pode ser caracterizado, em termos de coleta e análise de dados, como uma pesquisa de métodos mistos. Esta abordagem de investigação combina ou associa as formas qualitativa e quantitativa (CRESWELL, 2010). Sem deixar de observar a *indisciplina* de Triviños, haveremos de usar, em conjunto, as duas abordagens. Enquanto coleta, preparação e processamento de dados, o trabalho se fundamenta em métodos puramente quantitativos. Mas a abstração do processo de retenção, a escolha do modelo, a discussão e a análise dos achados, far-se-ão valer dos métodos qualitativos.

Em termos gerais a pesquisa é concebida nas seguintes etapas:

1. Construção do banco de dados a partir do CENSUP 2020 e posterior enriquecimento da base com dados provenientes do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) da UNEMAT;
2. Tratamento e limpeza dos dados;

3. Seleção de *features* (variáveis) alinhadas ao modelo teórico de retenção adotado;

4. Clusterização dos Dados

5. Análise e Apresentação dos Resultados.

Na etapa 1, a qualidade da base de dados na qual o algoritmo de clusterização operará é diretamente proporcional à qualidade do resultado final. Nesta proposta, extraímos os dados da UNEMAT disponíveis no CENSUP 2020<sup>4</sup> criando nossa primeira estrutura. O CENSUP traz aproximadamente 100 variáveis para cada estudante ali registrado. As variáveis refletem diversas características dos estudantes. O censo também registra dados da instituição, dos cursos e do corpo docente. Do nosso interesse foram as variáveis dos estudantes, mas acessamos também os microdados dos cursos para obter o tempo mínimo de integralização deles, assim como os locais onde são ofertados. Então filtramos todos os registros referentes à UNEMAT, recuperando um total de 21.786 estudantes. Nesta primeira fase já foi possível uma primeira análise exploratória dos dados e alguns indicadores gerais da instituição foram extraídos como: taxas de diplomação, de retenção e de evasão, tempo médio de duração dos cursos e tempo médio de vínculo com a instituição dos estudantes diplomados.

Pelos microdados dos cursos contidos no censo, identificamos a duração mínima dos cursos de todos os estudantes do nosso banco de dados. Neste ponto nosso interesse era identificar quais estavam na condição de retidos. Disso, filtramos aqueles estudantes com tempo mínimo de integralização já transcorrido e sem diplomação. Nosso filtro retornou 2.455 estudantes.

Constituído, então, nosso banco de dados de estudantes retidos, seguimos enriquecendo-o com o coeficiente de rendimento acadêmico (CRA) dos estudantes. Isso foi feito por meio de consultas ao SIGAA da UNEMAT. A interface de consultas do SIGAA não nos permitiu obter individualmente o CRA, de modo que categorizamos em A, B, C aqueles estudantes com CRA, respectivamente, abaixo de 5, igual a 5 e inferior a 7, e acima de 7 pontos. Levamos em conta a normativa acadêmica da instituição que instrui que estudantes, numa dada disciplina, com média final abaixo de 5 pontos, são estudantes reprovados; com média final entre 5

---

4 Até o início desse projeto era o último censo disponibilizado

e 7 pontos, são estudantes em Prova Final e, média acima de 7 pontos, são estudantes aprovados.

A segunda etapa, de limpeza e tratamento de dados tem como objetivo uniformizar os dados, descartar registros incompletos, discretizar as variáveis contínuas.

A terceira etapa é a escolha das variáveis (*features*). Salientando que do conjunto de variáveis presentes no nosso banco de dados, selecionamos aquelas que, segundo o modelo teórico de Vincent Tinto, poderiam refletir algum aspecto relacionado à retenção. O quadro 1 descreve as variáveis que foram utilizadas na pesquisa.

Aqui caberia dizer que nosso processo de pesquisa fora, em certos aspectos, prejudicado devido à dificuldade no acesso aos dados dos sistemas institucionais. Apesar de diversas solicitações formais, enfrentamos entraves burocráticos.

**Quadro 1** - Descrição das Variáveis

VARIÁVEL	DESCRIÇÃO
RACA	raça declarada do estudante
IDADE	idade em anos do estudante
CURSO_CINE <sup>5</sup>	código CINE do curso do estudante
CRA	índice de rendimento acadêmico do estudante
CURSO_GRAU	descreve se o curso é bacharelado ou licenciatura
LOCAL_OFERTA	nome do local onde o curso é ofertado
STATUS	descreve se o estudante está trancado, transferido ou cursando
TURNO	o turno do curso
INGRESSO	o ano e o semestre de ingresso do estudante
NOME_CURSO	o nome do curso
TIPO_EM	tipo ensino médio: escola pública ou privada
NACIONALIDADE	nacionalidade dos estudante
FORMA_INGRESSO	tipo de ingresso: por vestibular, enem ou outra forma
ESTCIVIL	estado civil do estudante
SEXO	sexo do estudante
PERCENTCH	relação entre carga horária cumprida e carga horária total do curso

**Fonte:** Elaborada pelo Autor

Reconhecemos a natureza sensível dos dados e evidentemente o cuidado que a UNEMAT deve ter enquanto tutora dessas informações. Mas a demora nas

5 A Classificação Internacional Normalizada da Educação Adaptada para Cursos de Graduação e Sequenciais de Formação Específica do Brasil (Cine Brasil) é baseada na *International Standard Classification of Education – Fields of Education and Training (ISCED-F)* da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), composta por quatro níveis de classificação. A metodologia permite o agrupamento de cursos de educação superior segundo a semelhança de conteúdo temático, princípio básico do processo de classificação desses cursos. <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/cine-brasil>>

respostas e a incompletude que fora dada às nossas demandas, nos impediu que criássemos um modelo que utilizasse todas as fontes que razoavelmente nos eram possíveis. No entanto, as informações presentes no CENSUP garantiram uma quantidade suficiente de variáveis que possibilitaram capturar as diferentes dimensões dos estudantes que eram de nosso interesse e, conseqüentemente, à razoabilidade da nossa proposta.

A quarta etapa é a clusterização propriamente dita. Definimos para este trabalho o algoritmo *K-modes*. A seção seguinte 2.1 detalha essa técnica.

Por fim, a quinta etapa, com os grupos formados, analisamos e apresentamos os resultados – o texto da seção 5 trata especificamente dessa etapa.

## 2.1 Clusterização

A técnica de clusterização também é conhecida como agrupamento. Essa técnica pertence aos métodos não supervisionados de aprendizagem de máquina. Referimo-nos à aprendizagem de máquina como sendo uma área interdisciplinar, interfaceando-se sobretudo com a análise estatística de dados, inteligência artificial, reconhecimento de padrões e visualização de dados. Seu objetivo é extrair informações de fontes volumosas de dados. São chamados de não supervisionados porque os dados observados não possuem nenhum rótulo que o identifica, cabendo ao método descobrir tal informação. Segundo Russel e Norvig (2013, pp. 711-712) a formação não supervisionada de agrupamentos consiste em distinguir várias categorias em uma coleção de objetos.

Algumas definições importantes para entender a técnica de agrupamento são descritas a seguir:

**Cluster:** também chamado de grupo, de agrupamento ou de aglomerado – Conjunto de elementos que compartilham um grau de similaridade entre si. Elementos de um mesmo grupo tendem a ser homogêneos entre si ao passo que heterogêneos em relação aos elementos de outro grupo.

**Centróide:** elemento central de um cluster. Este elemento é o “líder” do grupo, todos os elementos pertencem a um grupo possuem características similares às do centróide, e conseqüentemente aos dos demais elementos do grupo.

**Distância:** descreve quão semelhantes os pontos (elementos) de um cluster são. Quanto menor a distância entre os pontos, mais semelhantes eles são. Diferentes técnicas de clusterização, utilizam diferentes fórmulas para o cálculo de distância.

## 2.2 O Algoritmo *K-Modes*

O *K-Modes* é um algoritmo de agrupamento criado por Huang (1998). É uma proposta alternativa para análise de agrupamento quando se trabalha com variáveis categóricas, ou variáveis numéricas contínuas discretizadas. Ele é baseado no algoritmo de agrupamento *K-Means* (K-Média), usado quando as variáveis observadas são numéricas e de domínio contínuo. Dessa forma, por meio de fórmulas matemáticas, é possível calcular a distância (geralmente a distância euclideana) de cada ponto de um grupo ao seu respectivo centróide. Depois das distâncias calculadas, as médias aritméticas das distâncias são encontradas e esses valores passam a ser os valores das observações do centróide.

Dois diferenças marcam o *K-Modes*:

Primeiro, a distância dos pontos do grupo ao centróide é calculada por medidas de (di)similaridades (fórmulas 2.1 e 2.2) e;

Segundo, o centróide é atualizado com a moda (valor de maior frequência) das medidas de (di)similaridades.

O mecanismo de localização do centróide no *K-Modes* é semelhante ao do *K-Means*.

$$d_1(X,Y) = \sum_{i=1}^n \delta(x_i, y_i) \quad (2.1)$$

$$\delta(x_i, y_i) = \begin{cases} 0, & x=y \\ 1, & x \neq y \end{cases} \quad (2.2)$$

onde:

$x_i$  é o valor da  $i$ -ésima variável do dado  $X$ ,  $y_i$  é a o valor da  $i$ -ésima variável do dado  $Y$ ,  $n$  é o número de variáveis.

A abordagem não hierárquica de clusterização tem a característica de não definir a priori o número  $k$  de grupos. A depender dos próprios dados e da heterogeneidade deles é possível encontrar o número ideal que consegue separar bem os elementos. Ainda em Huang (1998), é proposto um método de determinação do número de grupos baseada numa função de custo (fórmula 2.3). O custo é a soma de todas as dissimilaridades entres os clusters.

$$WCD = \sum_{j=1}^k \sum_{i=1}^m d_1(x_i, y_c) \quad (2.3)$$

**Figura 1 - O Algoritmo K-Modes**

1. Selecione  $k$  observações e as use como centróides dos clusters;
2. Aloque cada observação no cluster mais próximo se baseando na medida de dissimilaridade. Atualize a moda de cada cluster após cada alocação;
3. Depois de todas as observações terem sido alocadas ao cluster, compare a dissimilaridade entre cada observação e a moda. Se for identificado que alguma observação está mais próxima de outro cluster, mova-a para o cluster mais apropriado e atualize a moda de ambos os clusters;
4. Repita o passo 3 até que nenhuma observação mude de cluster.

**Fonte:** Huang (1998)

Na fase de clusterização aplicamos o algoritmo *K-Modes* (figura 1) para a geração não supervisionada dos grupos de estudantes. A nossa função de

dissimilaridade é baseada na quantidade de variáveis em comum que os estudantes apresentam.

Em termos de ferramentas de softwares, neste projeto utilizamos o Gerenciador de Banco de Dados Neo4j<sup>6</sup> na versão 4.3.1. O Neo4j é um sistema de gerenciamento de banco de dados grafo desenvolvido pela Neo4j, Inc. Descrito por seus desenvolvedores como um banco de dados transacional compatível com ACID com armazenamento e processamento gráfico nativo. O Neo4j é implementado em Java e acessível a partir de software escrito em outras linguagens usando a linguagem de consulta Cypher. Esse gerenciador de banco de dados, por não necessitar de um esquema *a priori*<sup>7</sup>, otimizou nosso processo de leitura dos microdados do INEP que possuem milhares de registros. Como dito, essa ferramenta oferece também uma linguagem eficiente de consulta (Neo4j Cypher) que nos permitiu fazer a análise exploratória do banco de dados. A partir do banco de dados em um grafo, combinamos as suas funcionalidades às da linguagem Python<sup>8</sup>, versão 3.0, onde foi possível aplicar o algoritmo de clusterização.

---

6 <https://neo4j.com/v2/>

7 Não é necessário tornar explícito um esquema, como tabelas por exemplo, para carga dos dados. Esse esquema está implícito nas informações lidas e o banco de dados as organiza da forma mais conveniente

8 <https://www.python.org/>

### 3 REVISÃO DE LITERATURA

O trinômio permanência, retenção e evasão, na seara do ensino, embora distintos por definição, na prática estão diretamente interligados. Essa ligação fica mais evidente ao analisarmos as relações permanência/retenção e retenção/permanência. O fenômeno de evasão foi o primeiro a receber atenção de pesquisa e os achados foram subsidiando estudos dos demais processos de retenção e permanência.

Pelo balanço de produção, identificamos que o estudo do fenômeno da retenção aparece pouco na literatura como um objeto principal de estudo. Ele frequentemente é mencionado e discutido em trabalhos investigativos de Evasão e/ou de Permanência. Nesta perspectiva, não são raros os estudos de retenção no nível superior. De fato, após o trabalho da Comissão do MEC (ANDIFES, ABRUEM, SESU-MEC, 1996), um esforço considerável de pesquisa na área se constituiu no Brasil, em consonância com todos os países do mundo todo.

O levantamento foi feito no Catálogo de Teses e Dissertações da Capes<sup>9</sup>, recuperando numa primeira consulta 130 trabalhos. Os descritores de busca e filtros utilizados estão descritos no quadro 2.

**Quadro 2** - Primeiro Levantamento de Produções na área

<b>Veículo</b>	Banco de Teses e Dissertações da Capes
<b>Recorte temporal</b>	2015 a 2020
<b>Descritores</b>	“Retenção” AND “Ensino Superior” AND “Universidade”
<b>Grandes Áreas de Conhecimento</b>	Ciências Sociais Aplicadas, Humanas e Multidisciplinar
<b>Tipo de trabalho</b>	Dissertação de Mestrado e Tese de Doutorado

**Fonte:** Elaborado pelo Autor

Após a sistematização dos trabalhos, foi possível selecionar 21 deles que mais se correlacionam com nossa temática de pesquisa. O quadro 3 sumariza essas produções.

**Quadro 3** - Sumarização dos trabalhos correlacionados

9 <https://catalogodeteses.capes.gov.br/catalogo-teses/>

<b>ANO</b>	<b>AUTOR</b>	<b>TÍTULO</b>	<b>INSTITUIÇÃO</b>	<b>UF</b>
2015	ESTHER NUNES KLEIN GAMA	OBSTÁCULOS À FORMAÇÃO NO ENSINO SUPERIOR: UM ESTUDO DA RETENÇÃO DISCENTE NOS CURSOS DE GRADUAÇÃO PRESENCIAL DO CENTRO DE ARTES DA UFES	UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO	ES
2015	BRUNO CESAR DE NAZARETH CIRIBELLI	RETENÇÃO E EVASÃO ESCOLARES NO BACHARELADO INTERDISCIPLINAR EM CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA	UNIVERSIDADE FEDERAL DE JUIZ DE FORA	MG
2016	MARCO ANTONIO DA SILVEIRA SANTOS	FATORES DETERMINANTES DA EVASÃO NO ENSINO SUPERIOR	CENTRO UNIVERSITÁRIO DAS FACULDADES METROPOLITANAS UNIDAS	SP
2017	LUCI ANNEE VARGAS CARNEIRO	DESEMPENHO ACADÊMICO NO ENSINO SUPERIOR: UMA ANÁLISE CONTEXTUAL	UNIVERSIDADE FEDERAL DE SANTA MARIA	RS
2017	RITA DE CASSIA FINAMOR NICOLA	EVASÃO E RETENÇÃO DE ACADÊMICOS NA GRADUAÇÃO EM ENSINO SUPERIOR PRESENCIAL: O CASO DA URI CAMPUS SANTIAGO	UNIVERSIDADE DE SANTA CRUZ DO SUL	RS
2017	GIDEON SOARES DA SILVA	RETENÇÃO E EVASÃO NO ENSINO SUPERIOR NO CONTEXTO DA EXPANSÃO: O CASO DO CURSO DE ENGENHARIA DE ALIMENTOS DA UFPB	UNIVERSIDADE FEDERAL DA PARAÍBA	PB
2017	AMANDA CRISTINA DE SOUZA ALMEIDA	A IMPORTÂNCIA DO MARKETING DE RELACIONAMENTO NA RETENÇÃO DE ALUNOS NAS INSTITUIÇÕES PARTICULARES DE ENSINO SUPERIOR: UM ESTUDO DE CASO DE UMA INSTITUIÇÃO DE ENSINO SUPERIOR EM MG	UNIVERSIDADE METODISTA DE PIRACICABA	SP
2017	THIAGO LUIZ DE OLIVEIRA	FATORES DETERMINANTES DA RETENÇÃO DE ALUNOS: ESTUDO EM UMA INSTITUIÇÃO FEDERAL DE ENSINO	CENTRO UNIVERSITÁRIO ALVES FARIA	GO
2017	FRANCISCO COIMBRA CARNEIRO PEREIRA	MODELOS PREDITIVOS PARA EVASÃO DE ALUNOS NO ENSINO SUPERIOR PRIVADO - UMA APLICAÇÃO DE MACHINE LEARNING PARA GESTÃO DE MARKETING DE RELACIONAMENTO	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO	RJ
2018	JOEL DOS SANTOS BRANDAO	O IMPACTO DA EVASÃO E RETENÇÃO SOBRE O FINANCIAMENTO DE UNIVERSIDADES FEDERAIS BRASILEIRAS: UM ESTUDO A PARTIR DO INDICADOR ALUNO EQUIVALENTE	FUNDAÇÃO UNIVERSIDADE FEDERAL DO TOCANTINS	TO
2018	GUSTAVO SILVA GONCALVES	RETENÇÃO DE ALUNOS NO ENSINO SUPERIOR: VALIDAÇÃO DE ESCALA INTERNACIONAL E APLICAÇÃO DE MODELO PREDITIVO	UNIVERSIDADE NOVE DE JULHO	SP
2018	JULIANA OLIVEIRA BOMBARDELLI	FATORES DE PERMANÊNCIA DO INGRESSANTE NO ENSINO SUPERIOR EM UMA UNIVERSIDADE PRIVADA	UNIVERSIDADE DO VALE DO RIO DOS SINOS	RS

		COMUNITÁRIA DO RIO GRANDE DO SUL		
2018	JACQUELINE DE SOUSA COELHO	UM ESTUDO ACERCA DA RETENÇÃO DOS ALUNOS DO BACHARELADO INTERDISCIPLINAR EM ARTES E DESIGN E DEMAIS OPÇÕES DE SEGUNDO CICLO	UNIVERSIDADE FEDERAL DE JUIZ DE FORA	MG
2019	ALAN SANTOS DA SILVA	POLÍTICAS PÚBLICAS DE EDUCAÇÃO: JUBILAMENTO OU EXCLUSÃO? ESTUDO SOBRE EVASÃO E RETENÇÃO REALIZADO NA UNIVERSIDADE FEDERAL DO AMAPÁ	UNIVERSIDADE ESTADUAL DO CEARÁ	CE
2019	WELTON PINTO DE MIRANDA	EVASÃO E RETENÇÃO NO ENSINO SUPERIOR: O CASO DO CURSO DE CIÊNCIAS CONTÁBEIS DO CAMPUS SEDE DA UFJF	UNIVERSIDADE FEDERAL DE JUIZ DE FORA	MG
2019	NARA MILBRATH DE OLIVEIRA	A EVASÃO EM CURSOS SUPERIORES DE TECNOLOGIA: UMA ABORDAGEM BASEADA EM MODELAGEM PREDITIVA	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO GRANDE DO SUL	RS
2020	MARCIA NASCIMENTO HENRIQUES KNOP	RETENÇÃO E RESILIÊNCIA NO ENSINO SUPERIOR BRASILEIRO: DETERMINANTES DAS CHANCES DE CONCLUSÃO.	UNIVERSIDADE DE BRASÍLIA	DF
2020	MARIA APARECIDA MAES	MÉTODO DE APOIO A TOMADA DE DECISÃO ESTRATÉGICA NA RETENÇÃO DISCENTE NO ENSINO	UNIVERSIDADE FEDERAL DE SANTA CATARINA	SC
2020	JUCILENE MARIA DA SILVA DIAS	UM ESTUDO DA RETENÇÃO E DA EVASÃO NA SUSTENTABILIDADE DAS INSTITUIÇÕES DE ENSINO SUPERIOR E DA SUA GOVERNANÇA CORPORATIVA	CENTRO UNIVERSITÁRIO DAS FACULDADES METROPOLITANAS UNIDAS	SP
2020	CAMILA FERREIRA CAMPOS	CAUSAS DA EVASÃO E RETENÇÃO NO CURSO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO DA UNIVERSIDADE FEDERAL DA BAHIA	UNIVERSIDADE FEDERAL DA BAHIA	BA
2020	PAULO VICTOR DA SILVA VESCOVI	ANÁLISE PREDITIVA NA DETECTAÇÃO DE EVASÃO DE ALUNOS NO ENSINO SUPERIOR PRIVADO BRASILEIRO: ABORDAGEM DE ALGORÍTIMOS DE APRENDIZADO DE MÁQUINA COM BASE NAS PERSPECTIVA ACADÊMICAS, FINANCEIRAS, GEOGRÁFICAS E SOCIOECONÔMICAS	FUNDAÇÃO GETÚLIO VARGAS	SP

Fonte: Elaborado pelo autor a partir de dados extraídos do Catálogo de Dados e Teses e Dissertações da CAPES

Pela figura 2, a seguir, fica evidente que as pesquisas em retenção ainda se atrelam aos estudos de evasão e permanência, com maior ênfase no primeiro.

**Figura 2 - Termos recorrentes em trabalhos de retenção**



**Fonte:** Elaborada pelo autor

De forma geral, o que observamos é que retenção, ainda nos dias atuais, é tema de pesquisa em todo território nacional. Apresenta-se como uma questão tanto nas universidades públicas, quanto nas particulares. O ano 2016 concentrou a maior parte dos trabalhos, mas 2020 retoma uma possível tendência de crescimento.

Assim como podemos observar no trabalho de Nunes e Pereira (2019), nosso levantamento aponta que as produções acadêmicas acerca de retenção no Brasil constituem-se, majoritariamente, em estudos de caso e, em geral, as análises envolvem recortes de curso ou campus. Isso também é ilustrado pela figura 2. Neste norte, podemos apontar os trabalhos de Ciribelli (2015), Gama (2015), Soares da Silva (2017), Nicola (2017), Coelho (2018), Bombardelli (2018), Gonçalves (2018), Santos das Silva (2019), Miranda (2019) e Ferreira Campos (2020).

Não desconsiderando as contribuições desses últimos, damos um enfoque nos trabalhos de Knop (2020), Silva Dias (2020), Vescovi (2020), Maes (2020), Oliveira (2019), Brandao (2018), Pereira (2017), que mudam o foco da discussão ao considerarem questões tais como fatores socioeconômicos e retenção, gestão acadêmica, impactos financeiros da retenção nas universidades privadas, impactos no orçamento das universidades federais públicas, questões de modelos preditivos de retenção baseados em mineração de dados. Esses trabalhos são particularmente interessantes porque demonstram novas possibilidades de pesquisa e investigação.

A tese de Knop (2020) busca compreender as desigualdades internas do sistema de ensino superior brasileiro. A autora verifica a influência da origem social dos estudantes nas chances de conclusão dos cursos e avalia a distribuição dos graduandos entre as diferentes áreas de conhecimento e instituições de ensino superior. Nessa mesma perspectiva, Borges (2017) traça o perfil dos estudantes da UNEMAT, no campus de Cáceres, considerando os ingressos do SISU. A tese de Knop, parte da investigação do quanto as políticas de democratização têm contribuído para a equidade do sistema. Os resultados indicaram que a renda familiar continua sendo um fator muito importante para a conclusão do curso, tanto nas instituições públicas quanto nas privadas, quanto maior a renda, maior as chances de diplomar. A autora vai concluir que, de modo geral, apesar da expansão do acesso ao ensino superior, há uma expressiva estratificação horizontal nesse nível de ensino.

No trabalho de Silva Dias (2020) é feita uma análise de retenção como alternativa para conter evasão. Esse trabalho evidencia que a retenção afeta de igual modo as instituições privadas e as públicas. A autora analisa o funcionamento, as estratégias e o desenvolvimento de ações voltadas para retenção, focando no sucesso institucional e em questões financeiras nas instituições de ensino superior privada no país, além de atender aos indicadores de avaliação do Ministério da Educação. O texto chama atenção para o percentual alto de desistentes por meio de trancamentos e cancelamentos, impactando a receita das instituições. A pesquisa então propõe um Modelo de Gestão aplicável para redução da evasão, com a triangulação dos conceitos de sustentabilidade financeira, inovação e eficiência na gestão de modo a garantir a permanência acadêmica.

O estudo de caso apresentado por Vescovi (2020) apresenta um modelo empírico capaz de auxiliar a tomada de decisão gerencial quanto à previsão e à prevenção da evasão de estudantes em uma instituição de ensino superior privada. Ele analisa a utilização de variáveis acadêmicas, financeiras, geográficas e socioeconômicas aplicadas a quatro modelos supervisionados de aprendizado de máquina. Esse trabalho nos chama a atenção uma vez que demonstra a viabilidade da utilização de algoritmos de inteligência artificial como ferramenta de gestão pedagógica pelos responsáveis da retenção, assim chamados pelo autor. A

pesquisa demonstrou-se que a vantagem deste modelo é sua tempestividade e fácil interpretação dos resultados.

Maes (2020) apresenta uma rede bayesiana para apoiar a tomada de decisão estratégica na retenção discente em uma instituição de ensino superior privada, nos cursos na modalidade presencial, identificando variáveis que constituem fatores de risco à evasão. Tal qual em Vescovi, Maes fez valer de técnicas de *Machine Learning* (aprendizagem de máquina) e com dados da própria instituição, como sexo, situação financeira, benefícios/bolsas, situação acadêmica no momento da desistência, situação acadêmica anterior à desistência, forma de ingresso, turno frequentado, mobilidade, semestre do trancamento e quantidade de semestres frequentados até a evasão, o perfil dos estudantes evadidos fora retratado para apoio à gestão da instituição, auxiliando à tomada de decisão na prevenção e criação de políticas institucionais que permitam aos estudantes a finalização da graduação, diminuindo o impacto financeiro da instituição e adequando o planejamento docente e de infraestrutura institucional.

Oliveira (2019) pesquisa em um campus do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul, mais especificamente, em cursos da área de tecnologia da informação. Abordando modelos preditivos de evasão a proposta visa identificar, de maneira precoce, acadêmicos com propensão a abandonar o curso. Sendo assim, optou-se por usar técnicas de Mineração de Dados, dentro de um processo de Descoberta de Conhecimento em Banco de Dados (*KDD – Knowledge Discovery in Databases*). O trabalho utilizou dados acadêmicos e sociodemográficos, armazenados nos sistemas de acompanhamento acadêmico dos estudantes. Para construção do modelo de predição, foram avaliados cinco algoritmos de classificação. A pesquisa demonstra como pode ser facilitado o processo de tomada de decisão dos gestores da instituição, em relação às ações de permanência e êxito, a partir das predições indicativas de evasão.

Brandao (2018) mede a evasão e a retenção nos cursos de universidades federais entre os anos 2012 e 2016, verificando o efeito de algumas variáveis sócio demográficas e os seus impactos sobre o financiamento. Levando em conta que o indicador estudante equivalente, que ao padronizar seus valores ignora as peculiaridades dos cursos no que tange à localização geográfica e aos aspectos sociais, o autor usa os microdados do Inep e por meio de um modelo de regressão

estabelece a relação entre a evasão, a retenção e financiamento. Confirmou-se que o aumento da evasão e da retenção diminui o financiamento.

Pereira (2017) argumenta que o tratamento de retenção é uma das formas de se combater a evasão. Para a gestão de marketing de retenção, a identificação dos estudantes é o primeiro passo necessário para aplicar uma estratégia de interação personalizada. O trabalho apresenta uma metodologia quantitativa para classificação de risco de evasão de estudantes ativos. Baseado em dados históricos de estudantes que evadiram ou se formaram, modelos gerados por algoritmos de *machine learning* foram calculados e comparados e, na sequência, utilizados para classificar estudantes ativos.

No caso particular da UNEMAT, podemos destacar, como já citados, os trabalhos Souza (2020), de Hoffmann e Bitencourt (2019) e de Nodari, Lima e Maciel (2018).

Souza (2020) descreve as contribuições das Bolsas Acadêmicas ofertadas na graduação, UNEMAT – *campus* Cáceres, no período de 2019, para a permanência dos acadêmicos nos cursos da instituição. A autora vai concluir que na concepção dos bolsistas, vinculados aos Programas de Bolsas Acadêmicas da UNEMAT, apesar dos desafios e limitações inerentes aos programas, existem grandes contribuições para a permanência desses estudantes. As principais enumeradas pela autora são questões pedagógicas, acadêmicas e econômicas

Hoffmann e Bitencourt (2019) relacionam a evasão discente nos cursos de licenciatura da UNEMAT e as políticas de combate à evasão desenvolvidas pela Instituição, analisando o impacto dessas políticas no período de 2011 a 2015, mais especificamente as bolsas estudantis que evoluíram de 630 bolsas em 2011 para 4.220 em 2014. Utilizando dados do InepData fora constatado que a taxa média anual da evasão discente nas licenciaturas da UNEMAT, no período delimitado, oscilou entre 14% e 22%. Concluíram que as políticas desenvolvidas pela UNEMAT no período estudado não impactaram as taxas de evasão discente do mesmo período, pois as taxas de evasão oscilaram minimamente, em contraponto ao número de bolsas ofertadas, que pelas autoras são descritos como ‘carro chefe’ das políticas estudantis da UNEMAT.

Embora esses dois últimos trabalhos apresentem resultados bem distintos, é importante considerar a metodologia de cada um e o período de tempo investigado. De qualquer modo eles nos mostram como que os estudos de evasão e retenção são sensíveis aos sujeitos; ou a importância de estudar os impactos de políticas institucionais e ações qualificadores sobre evasão e retenção considerando as várias causas e motivos desses fenômenos.

Nodari, Lima e Maciel (2018) analisam a permanência a partir das correlações entre o desempenho dos estudantes no vestibular e durante a graduação. Partindo da questão se o desempenho dos estudantes no vestibular estaria correlacionado com a permanência destes nos cursos de graduação da UNEMAT – Campus/Cáceres, os autores fizeram uma investigação nos 12 cursos de graduação do campus de Cáceres. Os dados evidenciaram a idade média no ingresso (23 anos), o desempenho médio no vestibular (3,88), o desempenho médio na graduação (6,19). Em relação ao fluxo acadêmico - conclusão 27%, evasão 37%, retenção 32%, transferências 4%. Dos resultados, verificaram que maiores desempenhos no vestibular se associam a melhores desempenhos durante a graduação, influenciam a permanência e os índices de conclusão, evasão e retenção.

Um trabalho bem interessante que merece menção é o de Shuqfa e Harous (2019) que trazem uma revisão de literatura sobre Mineração de Dados Educacionais para explorar os métodos mais recentes usados na construção de modelos preditivos para retenção de estudantes e para prever as tendências futuras em diferentes contextos de ensino superior. Eles revisaram um conjunto diversificado de abordagens, modelos, conjuntos de dados, ferramentas, técnicas e medidas de desempenho. Os autores demonstram que as abordagens vão variar conforme o contexto educacional, onde oportunidades e desafios estão associados a cada abordagem.

Como descrito, os trabalhos focam em escopos mais restritos e no que diz respeito à mineração de dados educacionais, buscam modelos preditivos de retenção e evasão. Alguns ao trazerem variáveis descritoras preocupam-se numa descrição geral, como se fosse válido para todos os estudantes. Nossa proposta se assemelha em alguns aspectos aos trabalhos aqui descritos em partes, mas como diferencial é que estamos centrados na questão da retenção. Também abordamos

de forma mais generalizada nosso lócus, uma vez que consideramos todos os cursos de todos os campi da UNEMAT, e a aplicação de um método de clusterização que ainda não fora explorado conforme o fazemos neste projeto.

## 4 RETENÇÃO: Conceitos e Modelos Teóricos de Estudo

A Constituição Federal em seu Art. 208, inciso V descreve o dever do Estado em garantir acesso aos níveis mais elevados do ensino, da pesquisa e da criação artística, segundo a capacidade de cada um. O texto da lei nº 9.394, de 20 de dezembro de 1996 (LDBD), no inciso I seu do Art. 3º arrola a igualdade de condições para o acesso e permanência na escola como princípio basilar das condições do ensino. Ainda, a lei nº 13.005, de 25 de junho de 2014, que aprova o Plano Nacional de Educação – PNE, traça a meta 12 com o objetivo de elevar a taxa bruta de matrícula na educação superior para 50% (cinquenta por cento) e a taxa líquida para 33% (trinta e três por cento) da população de 18 (dezoito) a 24 (vinte e quatro) anos, assegurada a qualidade da oferta e expansão para, pelo menos, 40% (quarenta por cento) das novas matrículas, no segmento público.

Esse aparato legal é materializado em políticas públicas que pretendem dar conta da expansão do número de vagas e democratizar o acesso ao ensino superior. Mas esses mesmos instrumentos legais não explicitam políticas/ações para permanência desses estudantes, salvo os casos daqueles assistidos por cotas ou algum auxílio.

Cabe então à gestão universitária, por esforço próprio, gerenciar todas as questões de acompanhamento de permanência do seu corpo discente para que logrem êxito no fluxo ingresso - diplomação.

A evasão e a retenção são fenômenos que afetam esse fluxo. O primeiro implica redução das taxas de diplomação enquanto a retenção pode ter diferentes implicações.

Por ser um fenômeno multifacetado, o estudo de retenção no ensino superior, ao lado dos de ingresso, permanência e evasão, não é uma tarefa trivial. Pode-se querer identificar perfis de estudantes retidos, origens e causas de retenção, variáveis preditoras de retenção, tipos de retenção ou uma mescla de todas essas questões.

Ademais, não existe uma definição consensual de retenção. A literatura no Brasil e no exterior chega a oscilar entre extremos, onde de um lado a retenção é vista de forma prejudicial, ensejando fracasso estudantil, prejuízos financeiros, im-

pactos sociais negativos. E por outro lado, a retenção traz uma conotação positiva, podendo ser considerada sucesso nas políticas institucionais de permanência, fidelização do estudante-cliente nas instituições privadas de ensino superior, persistência estudantil (NUNES; PEREIRA, 2019).

Os primeiros estudos de retenção no ensino superior vão datar da década de 1930 e estiveram fundamentados na teoria do suicídio de Durkheim, do campo da sociologia, nas visões teóricas dos ritos de passagem nas sociedades tribais de Van Gennep no campo da antropologia social e do conceito de roteamento do trabalho do campo dos recursos humanos (COSTA; GOUVEIA, 2019). Esses autores destacam que na década de 1970, a introdução do termo “retenção” para descrever a persistência dos estudantes incluiu o conceito de que as instituições compartilhavam responsabilidade em influenciar as decisões dos discentes em relação ao abandono dos estudos.

Os modelos de retenção, devido a estreita ligação que este fenômeno tem com a evasão, são subsidiados fortemente pelos achados de pesquisa sobre o abandono. Costa e Gouveia (2019) permeiam as abordagens dos modelos de retenção de estudantes no ensino superior em seis perspectivas teóricas: a psicológica, a sociológica, a econômica, a organizacional, a interacional e a complementar ou integrativa, que compreende as perspectivas (sociológicos, econômicos, organizacionais e psicológicos).

A perspectiva psicológica se concentra nos atributos da personalidade do indivíduo. Em contrapartida, a perspectiva sociológica, não se concentra no individual, mas sim nas forças sociais que são externas a instituição educacional como status social, raça e prestígio. A perspectiva econômica se concentra no financiamento financeiro individual que afeta a retenção de estudantes. A perspectiva organizacional está em causa com o impacto de fatores organizacionais, como estrutura burocrática, tamanho e proporção de estudantes. A perspectiva interacional se concentra na influência da interação de indivíduos e fatores ambientais na retenção de estudantes. A perspectiva complementar ou integrativa compreende múltiplas perspectivas (sociológicos, econômicos, organizacionais e psicológicos).

Diferentes perspectivas levam a diferentes interpretações do fenômeno de retenção. A escolha do modelo, portanto, é uma etapa fundamental de qualquer pes-

quisa que investigue retenção: ele precisa abstrair este processo, captar os pontos mais relevantes e subsidiar as reflexões acerca dos achados da pesquisa. Por entendermos as múltiplas interfaces da retenção e por estarmos utilizando uma técnica de mineração de dados não supervisionada, entendemos que o modelo de Integração de Vincent Tinto seria um modelo interessante a ser adotado uma vez que trata as dimensões sociais, econômicas e acadêmicas dos estudantes numa perspectiva interacional.

#### **4.1 O Modelo de Retenção de Vincent Tinto**

O trabalho de Vincent Tinto tem acrescentado contribuições importantes para o estudo da evasão e retenção. Tinto é um dos maiores e mais conceituados estudiosos da retenção de estudantes no mundo (SEVERIANO SILVA, 2019). Ele sai de uma teoria mais generalista e com o tempo especializa sua teoria.

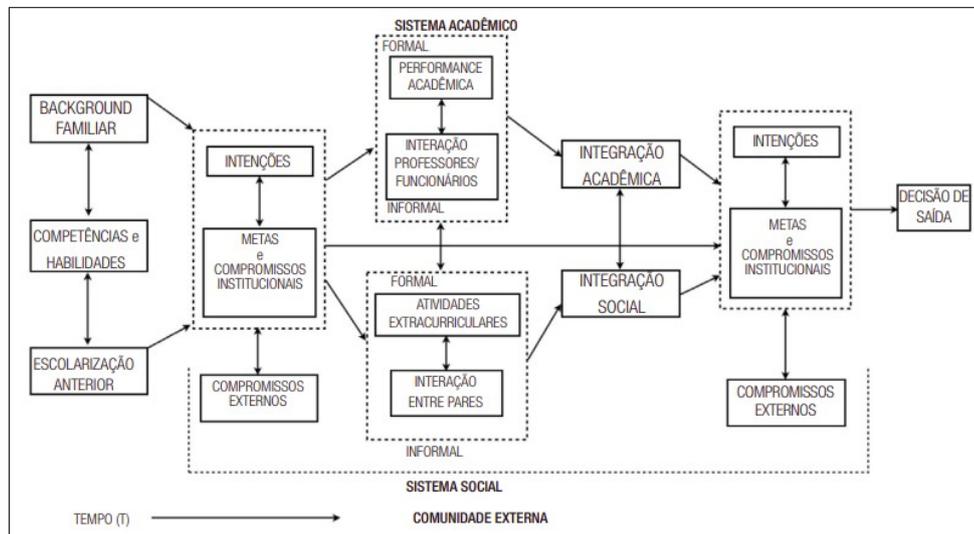
Seu trabalho seminal sobre o tema de evasão de estudantes data de 1975. Neste trabalho Tinto faz uma analogia da evasão com a teoria do suicídio de Durkheim, como se aquela fosse uma decisão que o separa da comunidade acadêmica. Posteriormente, lançou dois livros sobre o tema: *Leaving College* (Repensando as causas e curas da evasão estudantil) em 1983 e *Completing College* (Repensando a ação institucional) em 1993, com os quais aperfeiçoou o seu modelo de retenção.

Ele começa por um modelo integrador dos estudantes em seus contextos acadêmico e social – preconiza que um equilíbrio entre essas dimensões é condição para persistência do estudante na faculdade.

O modelo teórico de Tinto trabalha com variáveis interagindo entre si para determinar a decisão de evadir. A figura 3 esquematiza o modelo de interação. Quanto mais forte for o nível de integração acadêmica e social do estudante, maior o compromisso subsequente com a instituição e com a meta de obter a titulação. A evasão das instituições pode ser vista como oriunda de um processo longitudinal de interações entre um indivíduo com determinados atributos, habilidades, recursos financeiros, experiências de escolaridade prévia, disposições (intenções e

compromissos) e a integração dos membros dos sistemas acadêmico e social da instituição.

**Figura 3 - O Modelo de Integração de Tinto**



**Fonte:** Massi *et al* (2015)

O compromisso no momento de ingresso dos estudantes afeta a importância das interações acadêmicas e sociais dentro da instituição educacional e a intensidade de sua integração que, por vez, tem um impacto relevante nas suas metas e compromisso institucional.

Aqueles que entram na educação superior sem um forte propósito têm probabilidade de exibir baixos níveis de comprometimento assim como de persistência.

Tinto define também metas e compromissos que mostram a relevância das intenções dos estudantes e compromissos externos com a decisão de evadir ou desistir. Os compromissos externos com outros membros da sociedade e entidades fora da instituição, tais como: família, amigos e obrigações no trabalho têm um efeito contínuo e fundamental ao longo do tempo dispendido na faculdade.

As forças externas podem ser tanto de apoio quanto ter uma influência negativa nas metas e compromissos do estudante, subsequentes interações com a instituição e, finalmente, sua decisão de evadir. Espera-se que o grau de satisfação aumenta o grau de comprometimento pessoal com a instituição.

Ambas os escopos acadêmicos e não acadêmicos são vistos como tendo a capacidade de influenciar a decisão de evadir. Novamente, a comunidade externa, constituída por indivíduos e entidades com as quais o estudante interage, continua a ser um fator importante ao longo do tempo. Ele descreve que o arranjo de atividades sócio acadêmicas colaborativas gera um círculo virtuoso, pois a satisfação nas interações leva a um aumento na qualidade do esforço feito, pela alocação de tempo e energia, e leva a melhores resultados nas esferas social e acadêmica simultaneamente, o que reforça o sentimento de satisfação (TINTO, 1993).

Por considerar as diferentes dimensões do fenômeno de retenção, utilizamos nessa pesquisa o modelo de Tinto como fundamento teórico, tanto para a escolha das *features* na fase de clusterização quanto para embasar nossas discussões acerca do perfis apresentados.

## 5 RETENÇÃO NA UNEMAT

Uma primeira análise dos dados da UNEMAT contidas no censo de 2020 demonstrou alguns indicadores que discutimos a seguir.

O CENSUP 2020 revelou 21.786 estudantes dos cursos de graduação presenciais da UNEMAT, como sumarizado na tabela 1. O censo classifica esses estudantes como ‘*Cursando*’ que designa estudantes que ainda não cumpriram todas as exigências para obter seu grau; ‘*Matrícula Trancada*’ que designa estudantes com matrícula ativa mas sem cursar componentes curriculares daquele ano; ‘*Desvinculado do curso*’ para classificar estudantes que não realizaram matrícula e perderam o vínculo institucional; ‘*Transferido para outro curso da mesma IES*’ que designa estudantes com transferência acadêmica interna; ‘*Formados*’ que designa estudantes vinculados à instituição naquele ano que se formaram; e, ‘*Falecidos*’ para descrever estudantes que vieram a óbito.

**Tabela 1** – Situação dos estudantes da UNEMAT em 2019

Situação do estudante	Total de estudantes
Cursando	16354
Desvinculado do Curso	2501
Formado	2120
Matrícula Trancada	776
Transferido para outro curso da mesma IES	31
Falecido	4
<b>Total</b>	<b>21786</b>

Fonte: Elaborado pelo Autor. Baseado em Censup INEP (2020)

Neste trabalho consideramos que as taxas de retenção, diplomação e retenção seguem as fórmulas 5.1, 5.2 e 5.3 respectivamente<sup>10</sup>.

$$TaxadeReten\c{c}\tilde{a}o = \frac{EstudantesRetidos}{EstudantesRetidos + EstudantesDiplomados} \quad (5.1)$$

$$TaxadeDiploma\c{c}\tilde{a}o = \frac{EstudantesDiplomados}{EstudantesRetidos + EstudantesDiplomados} \quad (5.2)$$

$$TaxadeReten\c{c}\tilde{a}o = \frac{EstudantesDesvinculados}{TotalGeraldeEstudantes} \quad (5.3)$$

<sup>10</sup> Lima *et al* (2019) trazem uma discussão importante sobre taxas longitudinais de retenção e evasão. Os autores discutem as possibilidades e limitações dos indicadores ‘Taxa de Conclusão dos Cursos de Graduação (TCG)’ e a ‘Taxa de Sucesso na Graduação (TSG)’.

Dos 16.354 estudantes com status ‘Cursando’, foi possível filtrar, por meio da duração mínima de seus cursos<sup>11</sup>, aqueles estudantes que já haviam ultrapassado o tempo mínimo regulamentar de conclusão, portanto retidos na instituição. Identificamos neste ponto 2.455<sup>12</sup> deles. Isso nos dá um indicador de Taxa de Retenção de 53.66% e ao considerarmos os 2.120 formados naquele ano, uma Taxa de Diplomação de 46.34%. Especulamos que se essas retenções pudessem ser convertidas em diplomação, a instituição experimentaria um aumento considerável de estudantes diplomados.

A respeito dessa taxa de diplomação, as tabelas 2 e 3 mostram a evolução de 2009 a 2019<sup>13</sup> do número de vagas, ingressos, matrículas e concluintes no ensino superior no Brasil e na UNEMAT. Ainda que o número de matrículas englobe estudantes aptos e não aptos a concluírem, é possível estabelecer uma correlação entre os concluintes e matriculados para que tenhamos um comparativo entre o cenário nacional e o contexto da UNEMAT.

**Tabela 2 – Evolução das vagas, ingressos, matrículas e concluintes no ensino superior brasileiro**

<b>Ano Censo</b>	<b>Número de Vagas Oferecidas</b>	<b>Número de Tipo de Ingressos</b>	<b>Número de Matrículas</b>	<b>Número de Concluintes</b>
2009	3526933	1732613	5115896	826928
2010	3120192	1590650	5449120	829286
2011	3228671	1923927	5746762	865161
2012	3324407	2204456	5923838	876091
2013	4062474	2227545	6152405	829938
2014	5038392	2367401	6486171	837304
2015	5749175	2465751	6633545	916363
2016	6180251	2632500	6554283	938732
2017	6075252	2135126	6529681	947606
2019	6029702	2041136	6153560	934037

Fonte: Elaborado pelo Autor. Baseado em INEP Data, CENSUNP (2019;2020)

11 De acordo com o Censup 2020 os cursos da UNEMAT têm duração que varia de 3,5 a 6 anos (INEP 2020).

12 No momento de cruzarmos os dados do censo com os registros acadêmicos identificamos 286 registros com informações inconsistentes, de modo que os descartamos e para fins de clusterização trabalhamos com 2.169 alunos.

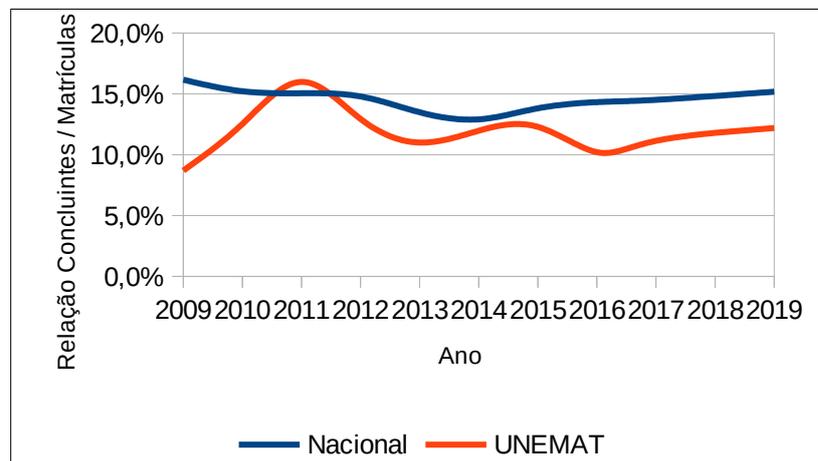
13 <<https://inepdata.inep.gov.br/>>

**Tabela 3** - Evolução das vagas, ingressos, matrículas e concluintes na UNEMAT

Ano Censo	Número de Vagas Oferecidas	Número de Tipo de Ingressos	Número de Matrículas	Número de Concluintes
2009	4400	4106	12163	1058
2010	4248	3192	13700	1716
2011	4290	3341	13554	2167
2012	5120	4897	14197	1837
2013	5340	3802	13671	1504
2014	5394	3828	13671	1637
2015	10620	4703	15869	1950
2016	5382	4708	16122	1646
2017	5329	4490	16870	1879
2019	5434	4826	17308	2120

Fonte: Elaborado pelo Autor. Baseado em INEP Data, CENSUNP (2019;2020)

Pela figura 4 identificamos que a instituição conseguiu um aumento expressivo dessa relação a partir de 2009, com queda em 2011 e mantendo-se abaixo do cenário nacional. Mais uma vez, a curva da UNEMAT poderia subir e razoavelmente ultrapassar a curva nacional, se de algum modo as retenções pudessem ser convertidas em diplomações aumentando o número de concluintes.

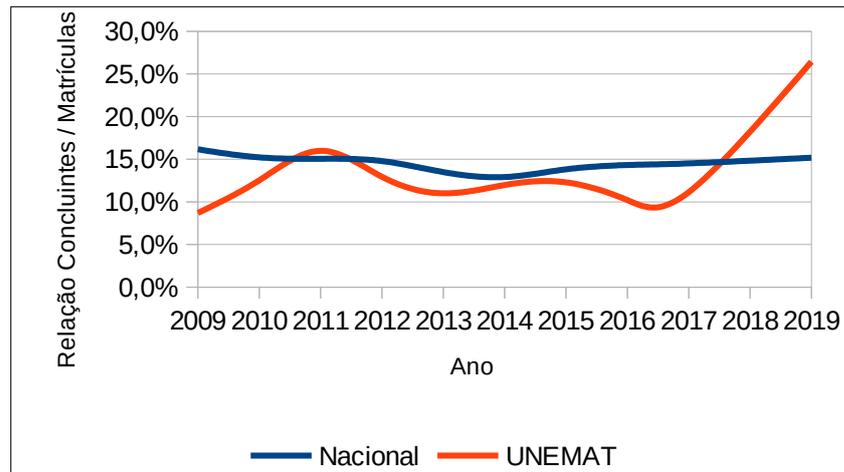
**Figura 4** – Relação Concluintes/ Matrículas

Fonte: Elaborada pelo Autor

A figura 5 ilustra o cenário hipotético, ao projetar os 2.455 retidos como concluintes, elevando o total destes para 4.575 em 2019. Na prática entendemos que não é tangível esse cenário, ele apenas ilustra potenciais melhorias nos índices da instituição. Outro ponto a ser considerado, como discutido em Brandão (2018),

são as questões relativas ao financiamento das instituições. Embora ali se discuta o caso da IFES, é plausível concebermos que as receitas de financiamento são proporcionalmente vinculadas à taxa de diplomação, ao passo que a evasão e a retenção são inversamente proporcionais à taxa de diplomação.

**Figura 5 –** Projeção Relação Concluintes/ Matrículas



**Fonte:** Elaborada pelo Autor

Ao voltarmos para os estudantes diplomados, identificamos que em média eles estiveram 5.57 anos na instituição. Se considerarmos que os cursos na instituição têm uma duração média de 4.5 anos, vamos deduzir ali também um possível perfil de estudante retido (esse número chama atenção para persistência/permanência). Adicionalmente, ao investigarmos os 2501 estudantes desvinculados, identificamos que 32.38% (810) deles estiveram vinculados à instituição por pelo menos 4 anos. Esse último montante revela que um percentual considerável de estudantes evadidos permaneceram vinculados à instituição tempo suficiente para integralizar quase que em totalidade seus cursos.

Para a taxa de evasão, os dados da instituição contidos no censo anterior ao ano estudado foi consultado. O censo 2019<sup>14</sup> revelou que 18.189 estudantes estiveram vinculados aos cursos de graduação presenciais da UNEMAT. Pelo censo 2020, identificamos que 2.501 destes estudantes haviam perdido seus vínculos. Isto nos apresenta uma taxa de evasão de 13.75%. Foi revelado, ainda, que em média estes estudantes cumpriram apenas 19.54% de seus créditos e quase a metade

<sup>14</sup> Dos 20537 estudantes de graduação presencial com vínculo ativo subtraímos 2344 que perderam vínculo e 4 óbitos.

deles, 47.82%, haviam ingressado na instituição depois de 2018. Observamos aqui o quanto a evasão incide nos semestres iniciais. A tabela 4 detalha o quantitativo de estudantes evadidos por ano de ingresso.

**Tabela 4** – Distribuição dos estudantes evadidos em função do ano de ingresso

<b>Ano de Ingresso</b>	<b>Estudantes Evadidos</b>
2008	3
2009	3
2010	11
2011	33
2012	42
2013	94
2014	149
2015	201
2016	274
2017	414
2018	829
2019	448
<b>Total</b>	<b>2501</b>

Fonte: Elaborado pelo Autor. Baseado em Censup INEP (2020)

O ano de ingresso de cada um dos 2455 estudantes retidos é apresentado na Tabela 5. A partir de 2011 já começamos a perceber uma quantidade mais significativa de estudantes nessa situação.

**Tabela 5** – Distribuição dos Retidos em função do ano de ingresso

<b>Ano de Ingresso</b>	<b>Estudantes Retidos</b>
2005	1
2006	3
2007	5
2008	14
2009	26
2010	69
2011	123
2012	224
2013	481
2014	831
2015	662
2016	16
<b>Total</b>	<b>2455</b>

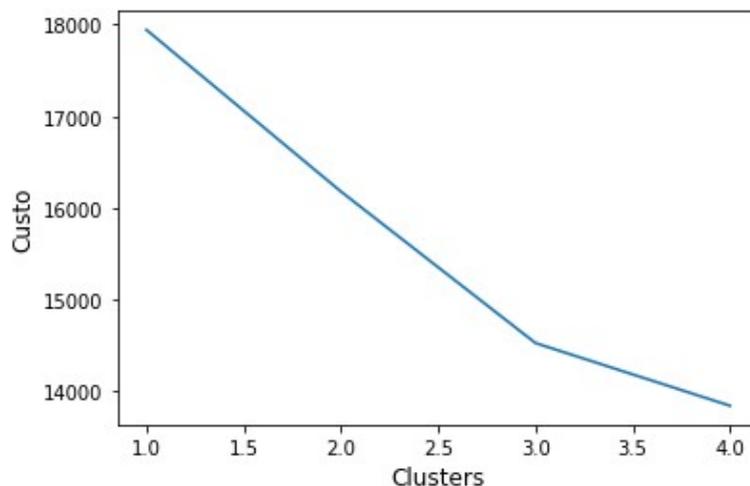
Fonte: Elaborado pelo Autor. Baseado em Censup INEP (2020)

Os dados discutidos até aqui dão conta do contexto geral da instituição. Com intuito de compreender como o fenômeno da retenção na UNEMAT é apresentado sob o óptica do perfil desses estudantes retidos, partimos para o processo de clusterização.

## 5.1 Clusterização dos Estudantes Retidos

Por se tratar de uma técnica não supervisionada, a quantidade  $k$  de clusters (agrupamentos) não é definida e/ou conhecida *a priori*. A definição da quantidade de grupos pode ser feita por meio de uma técnica chamada regra do joelho<sup>15</sup>. Esta técnica consiste em plotar o valor da função custo (fórmula 2.3) produzida por diferentes valores de  $k$ ; o valor de  $k$  onde o custo tem uma queda significativa é chamado de ‘joelho’ da curva, no qual indica que esse valor divide bem os dados. Pela figura 6, observamos que a partir de 3 grupos a curva de custo começa a ser suavizada, esse ponto (joelho da curva) sugere que este seja o número ideal de clusters já que a função custo (a soma total das diferenças entre os clusters) começa a diminuir.

**Figura 6** – Determinando o número de Clusters



Fonte: Elaborada pelo Autor

Considerando então 3 a quantidade ideal de clusters, aplicamos o algoritmo *K-Modes* (conforme seção 2.2) e como resultado, nosso conjunto de 2.169 registros foi particionado em três grupos, nomeados de cluster 1, cluster 2 e cluster 3, contendo 984, 715 e 470 estudantes respectivamente. O quadro 4 apresenta os centróides destes 3 clusters. Lembrando que cada centróide sumariza os valores mais frequentes (a moda) de cada variável do grupo. Pelos centróides, uma primeira caracterização dos grupos já pode ser feita.

<sup>15</sup> <https://www.oreilly.com/library/view/statistics-for-machine/9781788295758/c71ea970-0f3c-4973-8d3a-b09a7a6553c1.xhtml>

Ao analisarmos o comportamento dessas variáveis em cada um dos grupos conseguimos descrever o perfil dos estudantes retidos na instituição.

**Quadro 4 - Centróides dos Clusters**

	Cluster		
	1 (984)	2 (715)	3 (470)
<b>raca</b>	Parda	NA	Branca
<b>curso_cine</b>	0413A01	0114B01	0811A04
<b>curso_cine1</b>	4	1	8
<b>cra</b>	B	C	B
<b>grau</b>	1	2	1
<b>local</b>	Sinop	Cáceres	Tangará da Serra
<b>status</b>	1	1	1
<b>turno</b>	Noturno	Noturno	Integral
<b>curso_nome</b>	ADMINISTRAÇÃO	CIÊNCIAS BIOLÓGICAS	AGRONOMIA
<b>tipoem</b>	Pública	Pública	Pública
<b>tipoingresso</b>	ENEM	Vestibular	Vestibular
<b>estcivil</b>	Solteiro(a)	Solteiro(a)	Solteiro(a)
<b>sexo</b>	F	F	M
<b>faixaetaria</b>	20-25	20-25	20-25
<b>ingresso</b>	2015	2014	2014
<b>chpercent</b>	0.9-1	0.8-0.9	0.8-0.9

Fonte: Elaborado pelo autor

O grupo 1, o maior deles, contém 984 estudantes retidos. 58.43% desses estudantes estão nos cursos da grande área de Negócios, Administração e Direito (CINE 4) (Figura 7b). Os cursos de bacharelado representam uma porção considerável deste grupo (88.1%) (Figura 7m). Um pouco mais da metade dos estudantes do grupo 1 declaram-se pardos (50.30%) (Figura 7a). Eles concentram-se no campus de Sinop (24.08%), Tangará de Serra (16.36%) e Cáceres (13.21%) (Figura 7d). O turno noturno representa 62.40% desses estudantes (Figura 7e). A grande maioria desses estudantes são oriundos da escola pública (88.71%) (Figura 7g) e ingressaram na universidade via ENEM (58.84%) (Figura 7h). O ano de ingresso (Figura 7k) da maioria dos estudantes foi 2015 (48.17%). Em relação ao sexo não se tem uma diferença significativa, o grupo está igualmente distribuído entre ambos (Figura 7i). Pouco mais 16.16% desses estudantes não atingiram 50% de sua carga horária mínima de integralização de curso, ao passo que 52.44% deles

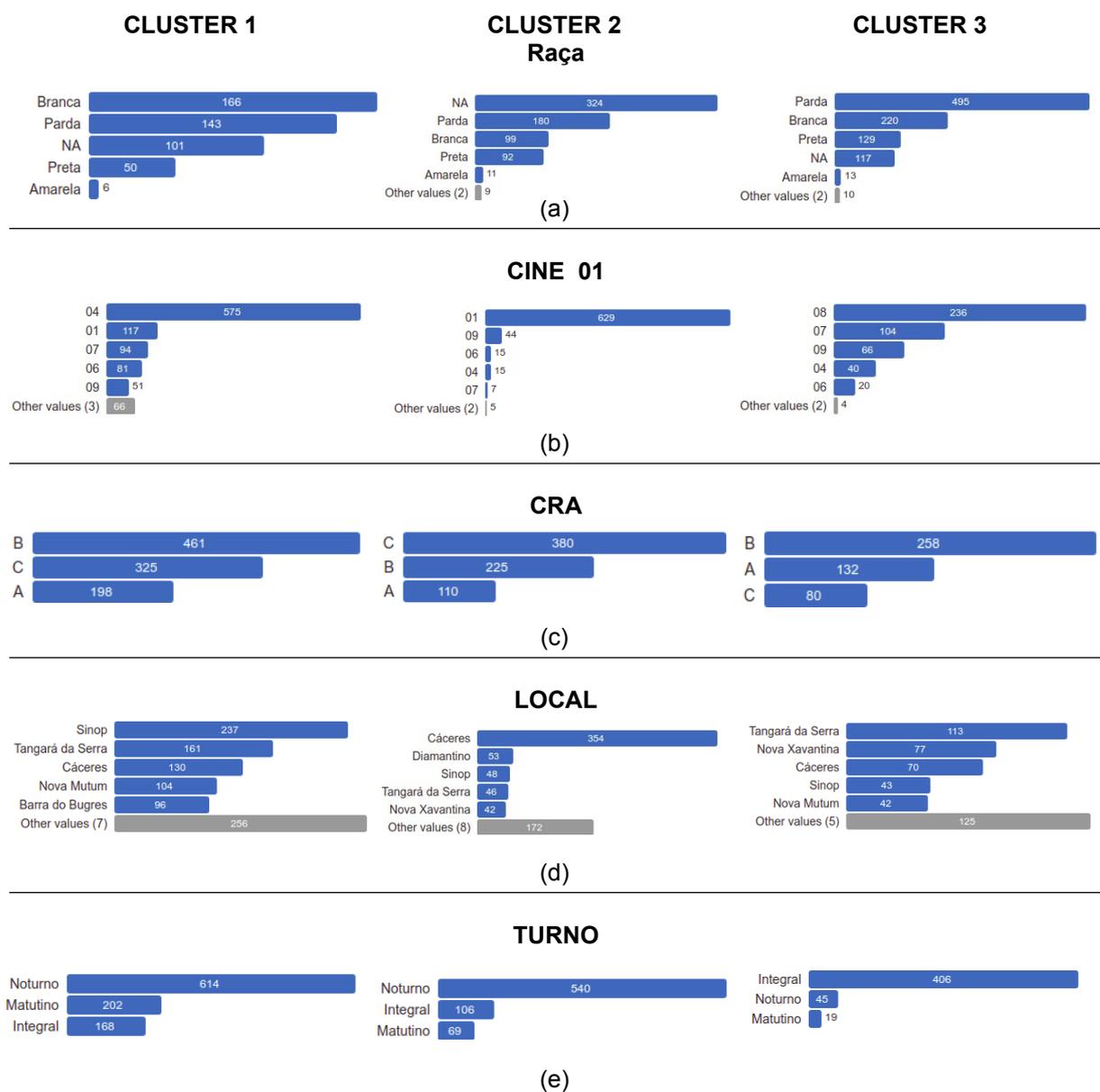
já têm mais 80% concluída (Figura 7l). Em termos de desempenho acadêmico (Figura 7c), 50.3% dos estudantes possui um coeficiente de rendimento maior ou igual a 5 e menor que 7. Isso indica que quase a metade desses estudantes tem se submetido a exames finais; 33.02% deles têm um CRA igual ou superior a 7, indicando aprovações sem exames finais e 20.12% desses estudantes têm experienciado reprovações.

O segundo maior grupo de estudantes retidos, o grupo 2 é formado por 715 estudantes. Aqui, 87.97% deles estão em cursos da grande área de Educação (CINE 1) (Figura 7b); As licenciaturas compreendem 88% destes estudantes (Figura 7m); 45.31% dos estudantes deste grupo não declaram suas raças (Figura 7a). Neste cluster, o campus de Cáceres (Figura 7d) é o que concentra a maior parte desses estudantes (49.51%); neste grupo predominam os cursos noturnos (75.52%) (Figura 7e). Quase que a totalidade são estudantes egressos de escola pública (93.85%) (Figura 7g). O tipo de ingresso mais frequente dos estudantes deste cluster é o vestibular (58.32%) (Figura 7h). Neste cluster o sexo feminino aparece com um pouco mais de frequência, 66.30% (Figura 7i). Os anos de ingresso desses estudantes concentram-se em 2014 e 2015 (60.7%) (Figura 7k). A carga horária integralizada (Figura 7l) de 34.27% dos estudantes do cluster 2 está entre 80 e 90%, 24.33% já concluíram mais de 90% de seus créditos e um montante de 15.24% ainda não atingiu mais que 50% de seus créditos. Sobre o desempenho acadêmico (Figura 7c), 53.14% apresentam um coeficiente de rendimento acadêmico igual ou superior a 7, o que indica aprovações sem exames finais. Com CRA no intervalo de 5 e 7, tem-se 31.46% dos estudantes e 15.3% deles apresentam um coeficiente de rendimento acadêmico inferior a 5.

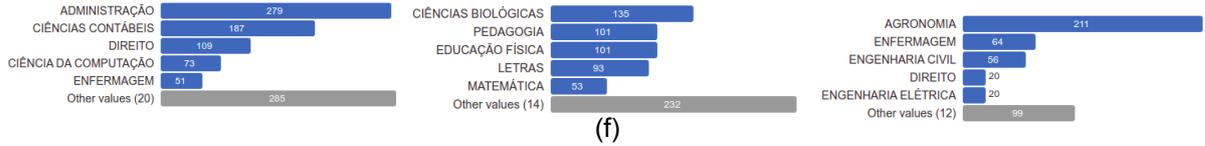
O grupo 3 compreende 470 estudantes. Desses, 35.32% declaram-se brancos e 30.42%, pardos (Figura 7a). Os cursos da grande área de Agricultura, Silvicultura, Pesca e Veterinária (CINE 8) (Figura 7a) concentram 50.21% dos estudantes enquanto que 22.13% dos estudantes encontram-se nos cursos da grande área de Engenharia, Produção e Construção (CINE 7). A figura 7f detalha esses cursos. Todos os discentes desse grupo estão em bacharelados (Figura 7m). A maioria deles estão no campus de Tangará de Serra (24.04%), seguido por 16.4% (Nova Xavantina) e Cáceres (14.9%) (Figura 7d). Estes estudantes estão essencialmente em cursos integrais (86.4%) (Figura 7e), também são egressos de

escola pública (80.85%) (Figura 7g) e 61.45% deles ingressaram via vestibular (Figura 7h). O sexo masculino é o mais frequente neste grupo (69.8%) (Figura 7i). 54.04% ingressaram no ano de 2014 (Figura 7k). A carga horária mínima de curso integralizada de 24.5% dos estudantes do cluster 3 está entre 80 e 90%, 23.83% já concluíram mais de 90% de seus créditos e um montante de 18.72% ainda não atingiram mais que 50% de seus créditos (Figura 7l). Sobre o desempenho acadêmico (Figura 7c), 54.9% dos estudantes têm um CRA no intervalo de 5 e 7, 28.08% deles possuem um CRA inferior a 5 e 17.02% dos estudantes apresentam um CRA igual ou superior a 7.

**Figura 7 – Distribuição de Frequência das Variáveis em cada cluster**



**CURSO**



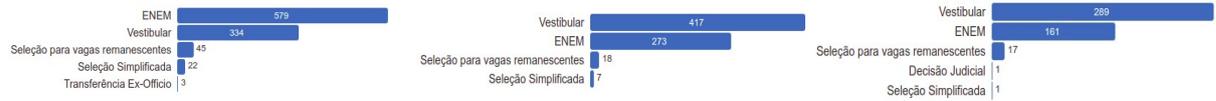
(f)

**ENSINO MÉDIO**



(g)

**TIPO DE INGRESSO**



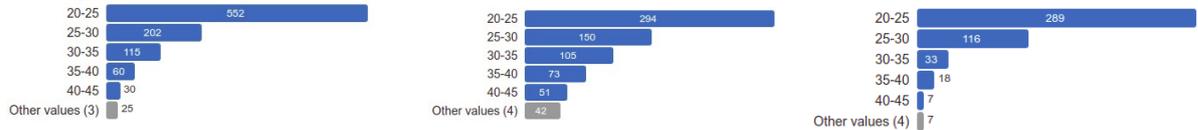
(h)

**SEXO**



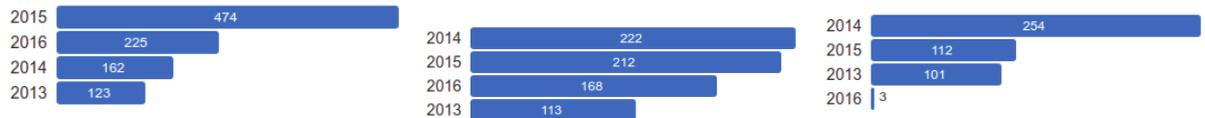
(i)

**FAIXA ETÁRIA**



(j)

**ANO DE INGRESSO**



(k)

**%CH CONCLUÍDO**



(l)

**CURSO\_GRAU**



(m)

Fonte: Elaborada pelo Autor

Reiteramos que a técnica de clusterização não nos permite explicar um fato. O método buscou agrupar os estudantes em função da maior similaridade entre eles. Esses grupos representam, em termos práticos, os perfis extraídos. De qualquer modo, o comportamento de algumas variáveis nos ajuda a entender e discutir esses perfis. Algumas apresentam uma distribuição de frequência bem particular enquanto outras já não se distribuem tão diferentemente.

As variáveis raça e faixa etária distribuem-se de forma semelhante nos três clusters. Elas revelam na verdade o perfil do estudante da UNEMAT e os clusters captam estas características: jovens com idade média de 25 anos e pardos.

A variável cine 01 é uma variável interessante, ela apresenta as maiores frequências no grupo 1 e 2. Disso, pode-se especular que hoje os estudantes retidos na UNEMAT concentram-se nos cursos da grande área de Negócios, Administração e Direito e da grande área de Educação Vescovi (2020) ao investigar métodos preditivos para evasão, ressalta que a variável qualitativa contendo o “nome do curso”, foi a de maior importância em todos os modelos investigados. O autor argumenta que a elevada importância desta variável provavelmente ocorre devido à diferença entre os perfis de estudantes que frequentam cursos distintos. Isto levanta uma outra questão acerca da democratização do ensino superior que é, segundo Knop (2020) sua estratificação. Em outros termos, indivíduos oriundos de estratos sociais mais elevados, com boa formação inicial tem maior probabilidade de ingressar em cursos de maior prestígio e retorno financeiro e maior chance de diplomação, os indivíduos do lado oposto, ingressam em cursos de menor retorno financeiro e estão mais propensos a evadir ou enfrentar dificuldades para conclusão.

A variável sexo caracteriza bem os grupos 2 e 3. No 2, em grande maioria estudantes de sexo feminino, enquanto no grupo 3, o sexo masculino é mais presente.

Pelo grau dos cursos observa-se que os grupos 1 e 3 são caracterizados pelos bacharelados e o grupo 2 pelas licenciaturas. Ressalta-se que 100% dos acadêmicos do grupo 3 são de bacharelados.

A variável turno já apresenta uma distribuição bem particular no grupo 3. Esse é o grupo de estudantes retidos que na maioria estão em cursos integrais.

A análise do CRA e do percentual de carga horária cumprida nos revela uma característica interessante dos estudantes retidos. Ela nos permitiu identificar o que Tinto chama de alta integração acadêmica. Os estudantes dos três grupos, em sua grande maioria, apresentam boas notas (CRA B e C), já cumpriram mais 80% da carga horária de seus cursos e de fato persistem na instituição. Corrobora com essa hipótese de persistência, o fato de que as diplomações têm ocorrido, em média, em 5,7 anos ao passo que a duração média dos cursos na UNEMAT é 4,5 anos. Em um de seus artigos mais recentes, Tinto traz uma discussão a respeito de retenção e persistência:

Durante anos a nossa visão predominante de permanência dos alunos tem sido moldada por teorias que veem a permanência dos alunos através das lentes de ação institucional e que perguntam o que as instituições podem fazer para reter seus alunos. Os estudantes, no entanto, não buscam serem retidos. Eles buscam persistir. (TINTO, 2017, p.254)

Por outro lado, 356 estudantes retidos (16.42%) ainda não atingiram 50% de suas cargas horárias. Uma investigação dos CRA's destes discentes nos revela que no grupo 1, 126 apresentam CRA inferior a 5, 26 estudantes com CRA entre 5 e 7 e 7 deles apresentam CRA superior a 7; no grupo 2, 85 tem CRA inferior a 5, 19 apresentam CRA entre 5 e 7, e 5 estudantes têm CRA superior a 7; no grupo 3, 79 estudantes têm CRA inferior a 5, 8 com CRA inferior a 7 e apenas 1 estudante tem CRA superior a 1. O que identificamos então é uma baixa integração acadêmica desses 356 indivíduos, uma vez que nos 3 grupos, essa parcela de estudantes está longe de integralizar seus créditos para diplomação, e ainda enfrentam problemas de notas. Pela Teoria de Tinto esses estudantes teriam grande probabilidade de evadirem.

Tinto destaca também a importância dos atributos pessoais do estudante antes de entrar na instituição de ensino superior. Embora não tenhamos adicionados todos os possíveis em no nosso modelo, os atributos Tipo de Ensino Médio e Turno poderiam nos dizer algo a respeito. Estudantes egressos de ensino médio de escola pública são predominantes em todos os clusters, no entanto, o grupo 1 e 2 a maioria dos estudantes estão em cursos noturnos o que pode implicar a necessidade de conciliar trabalho com estudos. Nesta mesma direção, Brandão (2018) também discute que o aumento da retenção para os cursos noturnos possui as mesmas

causas elencadas para a evasão: a necessidade de conciliar estudo com trabalho, a maior deficiência destes discentes, o menor envolvimento nas atividades extracurriculares da instituição de ensino, entre outras.

Feitas essas considerações, poderíamos então sumarizar os perfis desses estudantes como:

O grupo 1 é caracterizado por estudantes que se concentram nos bacharelados da área de Direito, Administração e Negócios; estão em média há 5,2 anos na universidade, já concluíram mais de 90% de sua carga-horária mínima, ingressaram via ENEM, estudam no turno noturno e possuem um CRA entre 5 e 7.

O grupo 2 é caracterizado por estudantes das licenciaturas da área de Educação, a maioria do sexo feminino, com CRA superior a 7, estão na universidade em média 5,4 anos, ingressaram via vestibular e estão no turno noturno com uma carga horária mínima cumprida entre 80 e 90%.

O grupo 3 é caracterizado por estudantes de cursos de período integral. Todos em bacharelados. Estão na instituição em média 5,95 anos, são na maioria de sexo masculino, ingressaram via vestibular, têm um CRA entre 5 e 7 e já cumpriram mais de 80% de sua carga horária mínima de integralização de curso.

Com os perfis constituídos e apresentados, um próximo passo, não alcançado pelo escopo deste trabalho, é uma investigação pontual e qualitativa sobre as reais causas de retenção considerando estes diferentes perfis.

## 6 CONSIDERAÇÕES FINAIS

Os desafios de estudar retenção, dadas a complexidade e a multidimensionalidade deste tema estiveram presentes neste trabalho. A falta de um consenso na conceituação do termo, as diferentes métricas de mensuração das taxas de retenção, o estreitamento desse fenômeno com os de permanência e evasão, tudo isso resume um pouco o que foram os desafios. Em termos operacionais, os desafios se configuraram na burocracia que nosso projeto enfrentou em acessar os dados institucionais e a não trivialidade de uso e manuseio dos microdados do censo da educação superior. Por força de lei, os dados são abertos e disponibilizados, mas, na prática, o acesso e uso dependem de computadores com razoável capacidade de armazenamento e processamento, conhecimentos de softwares específicos estáticos ou de banco de dados. Essas exigências podem impedir o uso efetivo desses dados.

Desafios a parte, o nosso trabalho ao trazer de forma inédita a técnica de clusterização, e mais especificamente pelo algoritmo *K-Modes*, para estudos de retenção no ensino superior se mostrou aplicável no estudo de caso da UNEMAT, podendo facilmente ser aplicada em qualquer instituição.

Pela análise exploratória dos clusters foi possível compreender o fenômeno de retenção nos cursos de graduação (modalidade presencial) da Universidade do Mato Grosso. Os 3 grupos gerados pela técnica de clusterização evidenciam que os estudantes retidos na UNEMAT têm uma alta integração acadêmica que implica expectativa de diplomação ainda que no prazo superior ao mínimo regulamentado nos PPC's. Mesmo nessa generalização, os perfis demonstram particularidades entre cada grupo.

Os atributos CRA, Percentual de Carga Horária Cumprida, Grande Área do Curso e Grau Acadêmico foram os que mais fortemente caracterizaram os clusters e trazem comportamentos bem distintos em cada agrupamento.

Adicionalmente, o banco de dados desenvolvido pode ser considerado como nossa contribuição tecnológica uma vez que subsidia a investigação de outros problemas de pesquisa como a evasão e permanência dos estudantes. O uso de

tecnologias diferentes das sugeridas pelo INEP mostra novas possibilidades de manipulação deste tipo de dado aberto.

Este trabalho trouxe uma possibilidade de uso de clusterização de dados educacionais. Como proposta para trabalhos futuros seria adicionar mais variáveis ao modelo e investigar o quão elas impactam na criação dos grupos. Explorar também novos algoritmos de clusterização, como aqueles de agrupamento hierárquicos também pode ser considerado. O trabalho analisou todos os cursos regulares presenciais da Universidade. Isso foi importante para que o contexto geral de retenção fosse entendido. Disso, podemos estender a proposta para uma dimensão mais reduzida, como por exemplo um curso ou um campus de modo a entender local e especificamente o fenômeno. Até mesmo os cursos da modalidade a distância podem ser estudados.

A pesquisa alcançou seu objetivo mas já pauta algumas especulações advindas da análise dos grupos: como o fenômeno de estratificação ocorre (e se ocorre) na instituição; como os cursos da área de Educação, em maioria noturnos, apresentam um elevado número de estudantes retidos ainda que com um alto coeficiente acadêmico? Teriam as diretrizes nacionais que determinam cargas-horárias mínimas dos bacharelados e das licenciaturas algum impacto sobre as diplomações além do prazo regulamentar? E na visão dos estudantes, eles se consideram retidos? Ou de fato persistem porque estão comprometidos com seus objetivos pessoais? Ou por atribuírem algum prestígio ao seu curso ou à Instituição?

Voltamos ao início, mas agora com um norte. Sabemos quem são os estudantes retidos, sabemos que seus perfis podem ser agrupados, sabemos as características mais relevantes desses perfis.

## REFERÊNCIAS

BETARELLI JUNIOR, Ademir Antonio. **Análise de Agrupamentos (Clusters)**. NOTA DE AULA DO PROGRAMA DE POS GRADUACAO EM ECONOMIA APLICADA. Juiz de Fora, 2016. Disponível em: <[https://www2.ufjf.br/lates//files/2016/12/Conte%C3%Bado-5-%E2%80%93-A\\_An%C3%A1lise-de-cluster-AA.pdf](https://www2.ufjf.br/lates//files/2016/12/Conte%C3%Bado-5-%E2%80%93-A_An%C3%A1lise-de-cluster-AA.pdf)>. Data de acesso: julho de 2020.

BORGES, Luiz Francisco. **PERFIL DOS INGRESSANTES NA UNIVERSIDADE DO ESTADO DE MATO GROSSO: IMPLICAÇÕES DO SISU NO PROCESSO DE DEMOCRATIZAÇÃO DO ACESSO**. Dissertação de Mestrado - Cáceres, MT: Universidade do Estado de Mato Grosso, Programa de Pós-Graduação em Educação, 2017.

BRANDAO, Joel dos Santos. **O IMPACTO DA EVASÃO E RETENÇÃO SOBRE O FINANCIAMENTO DE UNIVERSIDADES FEDERAIS BRASILEIRAS: UM ESTUDO A PARTIR DO INDICADOR ALUNO EQUIVALENTE**. Dissertação de Mestrado - Palmas, TO: FUNDAÇÃO UNIVERSIDADE FEDERAL DO TOCANTINS, 19 abr. 2018.

CAMPOS, Camila Ferreria. **CAUSAS DA EVASÃO E RETENÇÃO NO CURSO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO DA UNIVERSIDADE FEDERAL DA BAHIA**. Dissertação de Mestrado. Salvador, BA: UNIVERSIDADE FEDERAL DA BAHIA, 9 dez. 2020.

CIRIBELLI, Bruno Cesar de Nazareth. **RETENÇÃO E EVASÃO ESCOLARES NO BACHARELADO INTERDISCIPLINAR EM CIÊNCIAS EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA**. Dissertação de Mestrado. Juiz de Fora, MG: UNIVERSIDADE FEDERAL DE JUIZ DE FORA, 29 jun. 2015.

COELHO, Jacqueline de Sousa. **UM ESTUDO ACERCA DA RETENÇÃO DOS ALUNOS DO BACHARELADO INTERDISCIPLINAR EM ARTES E DESIGN E DEMAIS OPÇÕES DE SEGUNDO CICLO**. Dissertação de Mestrado - Juiz de Fora, MG: UNIVERSIDADE FEDERAL DE JUIZ DE FORA, 7 nov. 2018.

COMISSÃO ESPECIAL DE ESTUDOS SOBRE A EVASÃO NAS UNIVERSIDADES PÚBLICAS BRASILEIRAS ANDIFES/ABRUEM/SESU/MEC. **Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superior Públicas**. 1996.

COSTA, Oberdan Santos da; GOUVEIA, Luis Borges. MODELOS DE RETENÇÃO DE ESTUDANTES: ABORDAGENS E PERSPECTIVAS. In: **REAd. Revista Eletrônica de Administração (Porto Alegre)**, v. 24, n. 3, p. 155-182, set. 2018.

CRESWELL, John W. **Projeto de Pesquisa: Métodos Qualitativo, Quantitativo e Misto; tradução Magda Lopes**. 3 ed. Porto Alegre: Artmed, 2010.

HOFFMANN, Elaine; BITENCOURT, Loriege Pessoa. A EVASÃO DISCENTE NAS LICENCIATURAS DE MATEMÁTICA PRESENCIAIS DA UNEMAT (2011 A 2015) E AS POLÍTICAS DE COMBATE A ESSA EVASÃO. **Anais XIII Encontro Nacional de Educação Matemática**, ISSN 2178-034X, 2019.

GAMA, Esther Nunes Klein. **OBSTÁCULOS À FORMAÇÃO NO ENSINO SUPERIOR: UM ESTUDO DA RETENÇÃO DISCENTE NOS CURSOS DE GRADUAÇÃO PRESENCIAL DO CENTRO DE ARTES DA UFES**. Dissertação de Mestrado, Vitória, UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 9 dez. 2015.

GENTIL, Heloisa Salles. PESQUISA EDUCACIONAL: QUANTIDADE – QUALIDADE. **Revista Da Faculdade De Educação**, 15(1), 177–183. jun. 2011. Recuperado de <https://periodicos2.unemat.br/index.php/ppgedu/article/view/3890>

GONÇALVES, Gustavo Silva. **RETENÇÃO DE ALUNOS NO ENSINO SUPERIOR: VALIDAÇÃO DE ESCALA INTERNACIONAL E APLICAÇÃO DE MODELO PREDITIVO**. Dissertação de Mestrado - São Paulo, SP: UNIVERSIDADE NOVE DE JULHO, 19 dez. 2018.

HUANG, Zhexue. EXTENSIONS TO THE K-MEANS ALGORITHM FOR CLUSTERING LARGE DATA SETS WITH CATEGORICAL VALUES. **Data Mining and Knowledge Discovery**, v. 2, p. 283–304, 1998.

INEP. **CENSO DA EDUCAÇÃO SUPERIOR (2000-2019)**. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior>> Data de acesso: Março de 2020

KNOP, Marcia Nascimento Henriques. **RETENÇÃO E RESILIÊNCIA NO ENSINO SUPERIOR BRASILEIRO: DETERMINANTES DAS CHANCES DE CONCLUSÃO**. Tese de Doutorado - Brasília, DF: UNIVERSIDADE DE BRASÍLIA, 23 abr. 2020.

LÉVY, P. **Cibercultura** - Tradução de Carlos Irineu da Costa. 1a. ed. São Paulo: Editora 34 Ltda, 1999.

LIMA, Paulo *et al.* TAXAS LONGITUDINAIS DE RETENÇÃO E EVASÃO: UMA METODOLOGIA PARA ESTUDO DA TRAJETÓRIA DOS ESTUDANTES NA EDUCAÇÃO SUPERIOR. ***Ensaio: Avaliação e Políticas Públicas em Educação [online]***. 2019, v. 27, n. 102, pp. 157-178. Disponível em: <<https://doi.org/10.1590/S0104-40362018002701431>>. ISSN 1809-4465. Data de acesso: maio de 2020.

MAES, Maria Aparecida. **MÉTODO DE APOIO A TOMADA DE DECISÃO ESTRATÉGICA NA RETENÇÃO DISCENTE NO ENSINO**. Dissertação de Mestrado Florianópolis, SC: UNIVERSIDADE FEDERAL DE SANTA CATARINA, 18 dez. 2020.

MORIN, E. **Ciência com consciência** - Tradução Maria D. Alexandre.: Rio de Janeiro: Bertrand Brasil, 2010.

NICOLA, Rita de Cassia Finamor. **EVASÃO E RETENÇÃO DE ACADÊMICOS NA GRADUAÇÃO EM ENSINO SUPERIOR PRESENCIAL: O CASO DA URI CAMPUS SANTIAGO**. Dissertação de Mestrado - Santa Cruz do Sul: UNIVERSIDADE DE SANTA CRUZ DO SUL, 25 maio 2017.

NODARI, D. E.; LIMA, E. G. DOS S.; MACIEL, C. E. O DESEMPENHO DOS ESTUDANTES NO VESTIBULAR E A PERMANÊNCIA NOS CURSOS DE GRADUAÇÃO DA UNEMAT. ***Avaliação: Revista da Avaliação da Educação Superior (Campinas)***, v. 23, n. 2, p. 312–329, out. 2018.

NUNES, Sergio Inácio.; PEREIRA, Francielle Amâncio. **RETENÇÃO NO ENSINO SUPERIOR: REFLEXÕES A PARTIR DA PRODUÇÃO ACADÊMICA. *Anais VI CONEDU***. Campina Grande: Realize Editora, 2019. Disponível em: <<https://editorarealize.com.br/artigo/visualizar/62537>>. Acesso em: Junho de 2020.

OLIVEIRA, Nara Milbrath de. **A EVASÃO EM CURSOS SUPERIORES DE TECNOLOGIA: UMA ABORDAGEM BASEADA EM MODELAGEM PREDITIVA**. Dissertação de Mestrado - Porto Alegre, RS: INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO RIO GRANDE DO SUL, 19 dez. 2019.

PEREIRA, Francisco Coimbra Carneiro. **MODELOS PREDITIVOS PARA EVASÃO DE ALUNOS NO ENSINO SUPERIOR PRIVADO - UMA APLICAÇÃO DE MACHINE LEARNING PARA GESTÃO DE MARKETING DE RELACIONAMENTO**. Dissertação de Mestrado - Rio de Janeiro, RJ: PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO, 12 abr. 2017.

PEREIRA, Julio Cesar Rodrigues. **Análise de Dados Qualitativos: Estratégias Metodológicas Para as Ciências da Saúde, Humanas e Sociais**. 1a. ed. São Paulo: Edusp, 2004.

SANTOS DA SILVA, Alan. **POLÍTICAS PÚBLICAS DE EDUCAÇÃO: JUBILAMENTO OU EXCLUSÃO? ESTUDO SOBRE EVASÃO E RETENÇÃO REALIZADO NA UNIVERSIDADE FEDERAL DO AMAPÁ**. Dissertação de Mestrado. Fortaleza, CE: UNIVERSIDADE ESTADUAL DO CEARÁ, 16 maio 2019.

SHUQFA, Z.; HAROUS, S. DATA MINING TECHNIQUES USED IN PREDICTING STUDENT RETENTION IN HIGHER EDUCATION: A SURVEY. **INTERNATIONAL CONFERENCE ON ELECTRICAL AND COMPUTING TECHNOLOGIES AND APPLICATIONS (ICECTA)**, 2019, pp. 1-4, doi: 10.1109/icecta48151.2019.8959789.

SOARES DA SILVA, Gideon. **RETENÇÃO E EVASÃO NO ENSINO SUPERIOR NO CONTEXTO DA EXPANSÃO: O CASO DO CURSO DE ENGENHARIA DE ALIMENTOS DA UFPB**. Dissertação de Mestrado—João Pessoa, PB: UNIVERSIDADE FEDERAL DA PARAÍBA, 13 fev. 2017.

SOUZA, Lourdes Aparecida de. **PROGRAMAS DE BOLSAS NA UNEMAT: IMPLICAÇÕES PARA A PERMANÊNCIA**. Dissertação de Mestrado - Cáceres, MT: Universidade do Estado de Mato Grosso, Programa de Pós-graduação em Educação, 2020.

SILVA DIAS, Jucilene Maria da. **UM ESTUDO DA RETENÇÃO E DA EVASÃO NA SUSTENTABILIDADE DAS INSTITUIÇÕES DE ENSINO SUPERIOR E DA SUA GOVERNANÇA CORPORATIVA**. Dissertação de Mestrado, São Paulo, SP: CENTRO UNIVERSITÁRIO DAS FACULDADES METROPOLITANAS UNIDAS, 14 dez. 2020.

SILVA, M. S. **O MODELO DE RETENÇÃO DE ALUNOS DE VINCENT TINTO**. Disponível em <<https://www.linkedin.com/pulse/o-modelo-de-reten%C3%A7%C3%A3o-o-vincent-tinto-miro-severiano-silva>> . Data de acesso: Junho de 2020.

STUART Russel; PETER Norvig. **Inteligência Artificial**. 3. ed. São Paulo: Campus Elsevier, 2013.

TINTO, Vicente. Dropout from higher education: A theoretical synthesis of recent research. **REVIEW OF EDUCATIONAL RESEARCH**, v. 45, n. 1, p. 89–125, 1975.

TINTO, V. **Leaving College: Rethinking the Causes and Cures of Student Attrition**. London: University of Chicago Press, 1993.

TINTO, Vincent. THROUGH THE EYES OF STUDENTS. **JOURNAL OF COLLEGE STUDENT RETENTION: RESEARCH, THEORY & PRACTICE** 19 (2017): 254 - 269.

TRIVIÑOS, A. N. S. **Introdução à Pesquisa em Ciências Sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 1987.

WELTON PINTO DE MIRANDA. **EVASÃO E RETENÇÃO NO ENSINO SUPERIOR: O CASO DO CURSO DE CIÊNCIAS CONTÁBEIS DO CAMPUS SEDE DA UFJF**. Dissertação de Mestrado - Juiz de Fora, MG: UNIVERSIDADE FEDERAL DE JUIZ DE FORA, 18 out. 2019.

VESCOVI, Paulo Victor da Silva. **ANÁLISE PREDITIVA NA DETECTAÇÃO DE EVASÃO DE ALUNOS NO ENSINO SUPERIOR PRIVADO BRASILEIRO:**

**ABORDAGEM DE ALGORÍTIMOS DE APRENDIZADO DE MÁQUINA COM BASE NAS PERSPECTIVA ACADÊMICAS, FINANCEIRAS, GEOGRÁFICAS E SOCIOECONÔMICAS.** Dissertação de Mestrado - São Paulo, SP: FUNDAÇÃO GETÚLIO VARGAS, 2 mar. 2020.