



UNIVERSIDADE FEDERAL DO TOCANTINS  
CAMPUS DE PALMAS  
PROGRAMA DE PÓS-GRADUAÇÃO MESTRADO PROFISSIONAL  
EM PROPRIEDADE INTELECTUAL E TRANSFERÊNCIA DE TECNOLOGIA PARA A INOVAÇÃO



**LEANDRA CRISTINA CAVINA PIOVESAN SOARES**

**SOLUÇÕES TECNOLÓGICAS PARA O PROBLEMA DA EVASÃO  
UNIVERSITÁRIA, SOB A ÓPTICA DE FERRAMENTAS DE  
INTELIGÊNCIA ARTIFICIAL**

Palmas/TO  
2020

LEANDRA CRISTINA CAVINA PIOVESAN SOARES

**SOLUÇÕES TECNOLÓGICAS PARA O PROBLEMA DA EVASÃO  
UNIVERSITÁRIA, SOB A ÓPTICA DE FERRAMENTAS DE  
INTELIGÊNCIA ARTIFICIAL**

Dissertação apresentada ao Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para a Inovação. Foi avaliada para obtenção do título de Mestre em Propriedade Intelectual e Transferência de Tecnologia para a Inovação e aprovada em sua forma final pelo orientador e pela Banca Examinadora.

Orientado: Doutor Rafael Lima de Carvalho

Palmas/TO  
2020

**Dados Internacionais de Catalogação na Publicação (CIP)**  
**Sistema de Bibliotecas da Universidade Federal do Tocantins**

---

S676s Soares, Leandra Cristina Cavina Piovesan.  
SOLUÇÕES TECNOLÓGICAS PARA O PROBLEMA DA EVASÃO  
UNIVERSITÁRIA, SOB A ÓPTICA DE FERRAMENTAS DE INTELIGÊNCIA  
ARTIFICIAL. / Leandra Cristina Cavina Piovesan Soares. – Palmas, TO, 2020.  
115 f.

Dissertação (Mestrado Profissional) - Universidade Federal do Tocantins  
– Câmpus Universitário de Palmas - Curso de Pós-Graduação (Mestrado)  
Profissional em Propriedade Intelectual e Transferência de Tecnologia para  
Inovação, 2020.

Orientador: Rafael Lima de Carvalho

1. Evasão Escolar. 2. Mineração de Dados Educacionais. 3. Aprendizagem  
de Máquina. 4. Inovação em Gestão Educacional. I. Título

**CDD 346.8**

---

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer  
forma ou por qualquer meio deste documento é autorizado desde que citada a fonte.  
A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184  
do Código Penal.

**Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os  
dados fornecidos pelo(a) autor(a).**

# FOLHA DE APROVAÇÃO

LEANDRA CRISTINA CAVINA PIOVESAN SOARES

## SOLUÇÕES TECNOLÓGICAS PARA O PROBLEMA DA EVASÃO ESCOLAR, SOB A ÓPTICA DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL

Dissertação apresentada ao Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para a Inovação. Foi avaliada para obtenção do título de Mestre em Propriedade Intelectual e Transferência de Tecnologia para a Inovação e aprovada em sua forma final pelo orientador e pela Banca Examinadora.

Data de aprovação: 02/07/2020



Prof. (a) Dr (a). RAFAEL LIMA DE CARVALHO

Orientador e Presidente da Banca de Dissertação de Mestrado

Banca Examinadora

---

Prof. Dr. Warley Gramacho da Silva - UFT

---

Prof. Dr. Fabrício Barros Gonçalves - IFF

---

Profª. Dra. Maria das Graças Ferraz Bezerra - MPEG

Palmas/TO  
2020

## **AGRADECIMENTOS**

Este trabalho é fruto de um sonho que muitas vezes parece ser impossível, mas perante a Deus e pela nossa força de vontade os sonhos se tornam realidade.

Agradeço em especial a Deus por essa conquista, por ter me guiado e auxiliado pelos caminhos até chegar ao seu final. A minha família que em todo momento teve-se presente nas minhas angústias, medos e incertezas. Dedico essa conquista ao meu esposo Aldeir, a minha filha Marcelli e meu filho Gabriel, pela paciência e incentivos que foram me dado no decorrer desses anos. Além disso, a minha querida filha Marcelli que se engajou comigo na descoberta de novos conhecimentos, passando a conhecer o assunto com destreza. Todos esses motivos fizeram com que tivesse mais forças para lutar e continuar sempre firme em busca das minhas realizações.

Também dedico minha gratidão ao professor e orientador Rafael, que se dispôs sempre de forma serena os auxílios e orientações para a condução dos trabalhos. Agradeço também, aos amigos que foram conquistados no mestrado e que serão lembrados com muito carinho. A Bia, Jefferson, Cintia, Nélio, Mauricio, Wandson e Bebel, obrigado por tudo.

E por fim, ao Programa de Pós-graduação em Propriedade Intelectual e Transferência da Tecnologia para a Inovação (PROFNIT) e a Universidade Federal do Tocantins (UFT), pela oportunidade de desenvolver novos conhecimentos a partir deste programa de mestrado.

## RESUMO

Esta pesquisa amplia o debate sobre a problemática da Evasão Escolar no Ensino Superior Brasileiro. Os prejuízos gerados pela interrupção no ciclo dos estudos são inúmeros e a sua consequência é no desenvolvimento humano, causando cada vez mais incompatibilidades socioeconômicas e culturais no País. Neste contexto, é imprescindível o desenvolvimento de métodos, processos e soluções inovadoras, capazes de auxiliar as Instituições de Ensino no combate à evasão. Assim sendo, o presente trabalho objetiva-se, de modo geral, conhecer o contexto da evasão escolar no ensino superior brasileiro; mapear e discutir os avanços tecnológicos no campo da Propriedade Intelectual e Propriedade Industrial; apresentar a aplicação das técnicas de inteligência artificial em um contexto acadêmico. E por fim, o desenvolvimento de uma solução inovadora e inteligente, capaz de prever os alunos com propensão de evasão e/ou graduar. Como material investigativo, explorou-se os dados acadêmicos dos cursos presenciais da Universidade Estadual do Tocantins (Unitins) e utilizou-se de bases de pesquisas, como teses, dissertações e artigos científicos. O estudo demonstrou que a Evasão Escolar é um fenômeno complexo, uma vez que este problema não está restrito às Instituições de Ensino Brasileiras, fenômeno este que aflige também o cenário internacional. Nota-se, que diversos estudos estão sendo desenvolvidos por pesquisadores, contribuindo cada vez mais com soluções e meios para contornar o problema. A pesquisa também mapeou nas bases tecnológicas tanto no Brasil e no mundo os quantitativos de soluções tecnológicas referentes a patentes e registros de programas de computador. Pelos números, percebe-se um quantitativo muito baixo, com poucos registros e, além disso, muitas dessas tecnologias foram anuladas ou extintas. Em se tratando do uso da Mineração de Dados Acadêmicas, para a descoberta do conhecimento sobre os dados da evasão escolar, a pesquisa utilizou os dados acadêmicos da Unitins, referente aos 12 cursos presenciais. As áreas utilizadas da inteligência artificial foram a Mineração de Dados Educacionais (MDE), Aprendizado de Máquina (AM) com o uso do algoritmo de floresta aleatória. A proposta metodológica permitiu a identificação proativa e acurada dos alunos com risco de evasão ou com predisposição a formar. O Sistema Inteligente tem como objetivo a disponibilização de informações sobre a evasão escolar de forma fácil, desde os primeiros períodos de curso. Com a consistência dos resultados, permitiu que a Universidade do Estado do Tocantins (Unitins), obtivesse 95,21% de precisão entre os períodos, certificando assim, a confiabilidade, a precisão, a robustez do algoritmo de floresta aleatória. Proporcionando a gestão a tomada de decisão de forma proativa e assertiva, antecipando-se para a reversão do quadro de evasão.

**PALAVRAS-CHAVE:** Evasão Escolar. Mineração de Dados Educacionais. Aprendizagem de Máquina. Inovação em Gestão Educacional.

## ABSTRACT

This research expands the debate on the issue of school dropout in Brazilian Higher Education. The losses generated by the interruption in the cycle of studies are numerous and their consequence is in human development, causing more and more socioeconomic and cultural incompatibilities in the country. In this context, it is essential to develop innovative methods, processes and solutions, capable of helping people. Educational Institutions in the fight against dropout. Therefore, the present work aims, in general, to know the context of school dropout in Brazilian higher education; map and discuss technological advances in the field of Intellectual Property and Industrial Property; to present the application of artificial intelligence techniques in an academic context. And finally, the development of an innovative and intelligent solution, capable of predicting students with a tendency to drop out and / or graduate. As investigative material, the academic data from the face-to-face courses at the State University of Tocantins (Unitins) were explored and research bases were used, such as theses, dissertations and scientific articles. The study demonstrated that school dropout is a complex phenomenon, since this problem is not restricted to Brazilian Educational Institutions, a phenomenon that also afflicts the international scenario. It is noted that several studies are being developed by researchers, increasingly contributing with solutions and means to overcome the problem. The research also mapped on the technological bases both in Brazil and in the world the quantitative of technological solutions referring to patents and registrations of computer programs. By the numbers, we see a very low quantity, with few records and, in addition, many of these technologies have been canceled or extinguished. When it comes to the use of Academic Data Mining, for the discovery of knowledge about school dropout data, the research used Unitins' academic data for the 12 on-site courses. The areas used for artificial intelligence were Educational Data Mining (MDE), Machine Learning (AM) using the random forest algorithm. The methodological proposal allowed the proactive and accurate identification of students at risk of dropping out or with a predisposition to graduate. The Intelligent System aims to provide information about school dropout easily, from the first periods of the course. With the consistency of the results, it allowed the University of the State of Tocantins (Unitins) to obtain 95.21% accuracy between the periods, thus certifying the reliability, precision, robustness of the random forest algorithm. Providing management with proactive and assertive decision making, anticipating the reversal of the evasion situation.

**Key-words:** School dropout. Educational Data Mining. Machine learning. Innovation in Educational Management.





## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> - Áreas de relacionamento com a Inteligência Artificial.....	27
<b>Figura 2</b> - Taxonomia das principais subáreas da pesquisa em MDE.....	31
<b>Figura 3</b> - Hierarquia do Aprendizado de Máquina.....	34
<b>Figura 4</b> - Modelagem dos Atributos.....	37
<b>Figura 5</b> - Quantitativos de alunos por <i>Status Acadêmico</i> .....	39
<b>Figura 5.1</b> - Principais etapas do processo de KDD.....	67
<b>Figura 5.2</b> - Quantitativo de alunos por tipo de status acadêmico.....	68
<b>Figura 5.3</b> - Matriz de confusão com as duas classes (Evadido e Não Evadido) .....	70
<b>Figura 5.4</b> - Curva AUC (A e B) .....	72
<b>Figura 6.1</b> -Etapas do processo de KDD.....	87
<b>Figura 6.2</b> -Quantitativo de alunos por tipo de status acadêmico.....	87
<b>Figura 6.3</b> - Matriz de confusão com as duas classes (Evadido e Não Evadido) .....	93
<b>Figura 6.4</b> - Curva AUC (A e B) .....	94
<b>Gráfico 4.1</b> - Quantitativo de Patentes por País.....	55
<b>Gráfico 4.2</b> - Países com patentes ativas.....	55
<b>Gráfico 4.3</b> - Depósito de Patentes por Ano.....	56
<b>Quadro 1</b> - Objetivos da CEUPB.....	21
<b>Quadro 2</b> - Distinção sobre conceitos de Evasão Escolar.....	22
<b>Quadro 3</b> - Temas do Congresso em 1956 sobre a Inteligência Artificial.....	23
<b>Quadro 4</b> - Abordagens sobre o conceito da Inteligência Artificial.....	28
<b>Quadro 5</b> - Descrição dos campos da tabela cadastro de alunos.....	38
<b>Quadro 6</b> - Descrição dos campos da tabela Matriz Curricular.....	40
<b>Quadro 7</b> - Descrição dos campos da tabela de Histórico Acadêmico.....	41
<b>Quadro 8</b> - Descrição dos tipos de pré-processamento.....	42
<b>Quadro 9</b> - Relação de cursos da Unitins por câmpus universitário.....	45

<b>Quadro 5.1</b> - Distinção sobre conceitos de Evasão Escolar.....	64
<b>Quadro 6.1</b> - Objetivos da CEUPB.....	83

## LISTA DE TABELAS

<b>Tabela 1</b> - Ingressantes da rede federal que mudaram de Unidade da Federação (UF) em 2018.....	17
<b>Tabela 4.1</b> - Expressões utilizadas nas buscas de Patentes e Programas de Computador.....	53
<b>Tabela 5.1</b> - Ranqueamento das disciplinas com maiores reprovações.....	65
<b>Tabela 5.2</b> - Resultados dos treinamentos.....	73
<b>Tabela 5.3</b> - Matriz de Confusão gerada a partir do treinamento.....	74
<b>Tabela 5.4</b> - Resultado dos classificadores dos grupos de 70% de treinamento e 30% para teste.....	74
<b>Tabela 5.5</b> - Resultado da Matriz de Confusão do grupo 70% de treinamento e 30% para teste.....	75
<b>Tabela 6.1</b> - Dados informativos dos ingressantes de 2010 com recorte em 2016.....	84
<b>Tabela 6.2</b> - Quantitativo de Atributos com valores nulos.....	88
<b>Tabela 6.3</b> - Atributos com maiores quantidades nulas.....	89
<b>Tabela 6.4</b> -Propensão de evasão por câmpus.....	95
<b>Tabela 6.5</b> -Ranqueamento dos 5 cursos com maiores índices de propensão de evasão.....	96
<b>Tabela 6.6</b> - Resultados dos treinamentos a partir do 1º período dos cursos presenciais.....	97
<b>Tabela 6.7</b> - Matriz de Confusão gerada a partir do treinamento dos períodos dos cursos.....	98
<b>Tabela 6.8</b> - Resultado dos classificadores dos grupos de 70% de treinamento e 30% para teste.....	99
<b>Tabela 6.9</b> - Matriz de Confusão do grupo 70% de treinamento e 30% para teste.....	100

## LISTA DE ABREVIATURAS E SIGLAS

AC	Acurácia
ABMES	Associação de Mantenedoras de Ensino Superior
AM	Aprendizado de Máquina
AUC	Área sob a Curva
BNet	BayesNET
CEP	Comitê de Ética em Pesquisa com Seres Humanos
CEUPB	Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras
EAD	Ensino a distância
EDM	Educational Data Mining
ENEM	Exame Nacional do Ensino Médio
FN	Falso Negativo
FP	Falso Positivo
FPR	Taxa Falso Positivo
IA	Inteligência Artificial
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
INPI	Instituto Nacional de Propriedade Industrial
KDD	Knowledge Discovery in Databases
MD	Mineração de Dados
MDE	Mineração de Dados Educacionais
MLP	MultilayerPerceptron
MU	Modelo de Utilidade
PCT	Tratado de Cooperação de Patentes
P&D	Pesquisa e Desenvolvimento
PRE	Precisão
PROFNIT	Programa de Pós-Graduação em Propriedade Intelectual e Transferência de Tecnologia para a Inovação
RA	Registro Acadêmico
REC	Revogação
RF	RandomForest
ROC	Características Operacionais do Receptor
SiMEU	Sistema de Modelagem de Evasão Universitária
SISU	Sistema de seleção simplificada
TI	Tecnologia da Informação
TRP	Taxa Verdadeiro Positivo
UF	Unidade Federativa
UFF	Universidade Federal Fluminense
UFPB	Universidade Federal da Paraíba
	UFPR Universidade Federal do Paraná
UFRN	Universidade Federal do Rio Grande do Norte
UFT	Universidade Federal do Tocantins
Unitins	Universidade Estadual do Tocantins
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

WEKA  
WIPO

Waikato Ambiente para análise do conhecimento  
World Intellectual Property Organization

## SUMÁRIO

1 INTRODUÇÃO	16
1.1 Justificativa	17
1.2 Objetivos	20
<b>1.2.1 Objetivo Geral</b>	20
<b>1.2.2 Objetivos Específicos</b>	20
1.3 Estruturas do Trabalho	20
2 REVISÃO DE LITERATURA	22
<b>2.1 A Evasão Escolar no Ensino Superior Brasileiro</b>	22
<b>2.2 Conceitos e a aplicações da Inteligência Artificial</b>	25
<b>2.3 O uso da Mineração de dados</b>	29
<b>2.4 A Tendência da Mineração de dados Educacionais</b>	31
3 METODOLOGIA	37
<b>3.1 Métodos e Ferramentas</b>	37
<b>3.2 A modelagem para a escolha dos melhores atributos</b>	38
<b>3.3 Delimitação do estudo: Amostra da pesquisa</b>	45
4 EVASÃO ESCOLAR COM BASE EM PATENTES E REGISTROS DE SOFTWARES	49
<b>4.1 Introdução</b>	49
<b>4.2 O cenário Nacional da Evasão Escolar</b>	50
<b>4.3 Sobre os Registros de Programas de Computador e Patentes</b>	51
<b>4.4 Metodologia</b>	54
<b>4.5 Resultados e Discussão</b>	55
<b>4.6 Considerações Finais</b>	57
<b>Referências</b>	58
5 APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA EM UM CONTEXTO ACADÊMICO COM FOCO NA IDENTIFICAÇÃO DOS ALUNOS EVADIDOS E NÃO EVADIDOS	63
<b>Introdução</b>	63
<b>O contexto da evasão escolar no ensino superior</b>	65
<b>Metodologia</b>	67
<b>Métodos de Aprendizado de Máquina</b>	69
<b>Métricas de avaliação da solução proposta</b>	70
<b>Resultados</b>	74
<b>Considerações Finais</b>	77

<b>Referências</b>	<b>78</b>
Sistema Inteligente para Previsão de Evasão Escolar em um Contexto Universitário	81
<b>6 Sistema Inteligente para Previsão de Evasão Escolar em um Contexto Universitário</b>	<b>82</b>
<b>INTRODUÇÃO</b>	<b>82</b>
<b>A EVASÃO ESCOLAR NO ENSINO SUPERIOR BRASILEIRO</b>	<b>84</b>
<b>MATERIAIS E MÉTODOS</b>	<b>87</b>
<b>OS RECURSOS DO APRENDIZADO DE MÁQUINA</b>	<b>91</b>
<b>AS MÉTRICAS DA AVALIAÇÃO DA SOLUÇÃO PROPOSTA</b>	<b>93</b>
<b>RESULTADOS</b>	<b>96</b>
<b>CONSIDERAÇÕES FINAIS</b>	<b>102</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>103</b>
7 CONSIDERAÇÕES FINAIS	107
7.1 Trabalhos Futuros	108
7.2 Produções Bibliográficas	108
<b>REFERÊNCIAS</b>	<b>110</b>
<b>APÊNDICES</b>	<b>116</b>
<b>Apêndice 1 - Comunicado de Registro do Programa de Computador do Sistema de Modelagem de Evasão Universitária (SiMEU - Web)</b>	<b>116</b>
<b>Apêndice 2 - Comunicado de Programa de Computador do Sistema de Modelagem de Evasão Universitária (SiMEU)</b>	<b>120</b>
<b>ANEXOS</b>	<b>125</b>
<b>Anexo 1 - Aprovação da Plataforma Brasil</b>	<b>125</b>

## 1 INTRODUÇÃO

A problemática da evasão escolar, ou abandono escolar são assuntos que são discutidos com frequência entre os meios acadêmicos (GUIMARÃES et al, 2019; GOMES et al., 2019; SABBATINI, 2015; SANTOS JUNIOR; REAL, 2019; ADACHI, 2017; CARVALHO, 2017; SOUZA, 2017). De acordo com (PEREIRA et al., 2011; BAGGI e colab., 2011) a evasão e a continuação por longo tempo dos estudos, são obstáculos para a Administração Pública que causam reflexos institucionais e sociais, por se tratarem de investimento de recursos humanos e financeiros. (FILHO e colab., 2007) analisa sob o ponto de vista institucional, que todos aqueles que abandonam seu curso trazem também questões financeiras. Segundo o autor este fator é o mais relevante e que explica a permanência do aluno na universidade.

Neste sentido, as Instituições de Ensino devem analisar, identificar e tomar ações, visando soluções para este problema.

Os estudos relacionados a evasão escolar, apresentam as motivações que levam o aluno a abandonar o seu curso. A evasão estudantil no ensino superior é um problema internacional que afeta o resultado dos sistemas educacionais. As perdas de estudantes que iniciaram, mas não terminaram seus cursos são desperdícios sociais, acadêmicos e econômicos. No setor público, são recursos públicos investidos sem o devido retorno. No setor privado, é uma importante perda de receitas. Em ambos os casos, a evasão é uma fonte de ociosidade de professores, funcionários, equipamentos e espaço físico (FILHO e colab., 2007).

Segundo GAIOSO (2005) o contexto da evasão escolar pode estar relacionado por falta de oportunidades de trabalho na área, laços afetivos com a instituição ou até mesmo sem um referencial familiar. Neste sentido, autores como (KIRA, 1998; GAIOSO, 2005; (BAGGI e colab., 2011) definem a evasão escolar como a interrupção do ciclo de estudos, em qualquer nível de ensino. TINTO (1975) justifica que não há uma definição capaz de compreender este fenômeno devido a sua complexidade, além de existir motivos com a falta de instrução sobre determinado curso, sobre a profissão ou sobre a própria estrutura curricular. Para (FILHO e colab., 2007) este problema é de ordem internacional e que afeta o desempenho dos sistemas educacionais.

Diante dos fatos apresentados, percebe-se o quanto este fenômeno é complexo, há vários aspectos envolvidos e que requer um estudo aprofundado. A necessidade de um monitoramento dos sistemas educacionais é justificada por meio dos dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), com alto grau de



desistência, permanência por longos tempos dos alunos nas instituições e pela média de anos de estudos dos Brasileiros (BRASIL, 1996c; INEP, 2019b). Estudar a evasão escolar no ensino superior a fim de obter informações e conhecimentos que possam ser utilizados para minimizá-lo é o objetivo deste trabalho. Assim, assume-se como hipótese a ser verificada que técnicas de Aprendizado de Máquina podem ser aplicadas de forma satisfatória em mineração de dados educacionais (MDE). Para a aplicação da hipótese utilizou-se como estudo de caso os dados acadêmicos da Universidade Estadual do Tocantins (Unitins), com a utilização de técnicas computacionais de aprendizado de máquina, onde os dados foram investigados e analisados a fim de prever a evasão antes de um aluno abandonar o curso.

O Aprendizado de Máquina (AM) é uma parte da Inteligência Artificial que compreende o desenvolvimento de técnicas com sistemas capazes de adquirirem conhecimento de maneira automatizada (MITCHEL, 1997). Desta forma, a partir de dados acadêmicos da Unitins, esta pesquisa aplicou o Aprendizado de Máquina para produzir um classificador preditivo, com o objetivo de aprender o padrão dos alunos evadidos e não evadidos. Ou seja, a partir dos primeiros períodos seja possível identificar precocemente os alunos propensos a evasão. Permitindo assim, uma ação proativa para minimizar o aluno em situação de abandono.

Este estudo torna-se significativo não apenas por subsidiar a análise de dados sobre a evasão, mas, além disso, por permitir que o modelo de Aprendizagem de Máquina e sua metodologia sirvam de base para outras questões relacionadas a dados acadêmicos, servindo como referência para outras instituições de ensino.

## **1.1 Justificativa**

A escolha deste tema ocorreu de forma a responder a uma demanda do setor de Tecnologia da Informação (TI) da Unitins. Ao observar os dados escolares desta instituição, percebeu-se uma necessidade por soluções inteligentes para amparar os gestores a respeito dos dados de evasão. Observou-se então que esta não era uma demanda isolada e que propor soluções nesta área, pode contribuir grandemente dentro da área de inovação no que tange à educação superior. Uma pesquisa preliminar fez-se perceber de que as maiorias das Universidades brasileiras necessitam de soluções para apoio à decisão que permitam embasar ações concretas contra a evasão. E pela busca deste entendimento, a Unitins desenvolveu em

2019, duas tentativas de contatos com os alunos evadidos, através de email e por telefone. Os resultados foram inconclusivos, pois a pesquisa não atingiu um número suficiente e representativo de alunos, para que fosse possível a construção de um diagnóstico.

Ao consultar os dados publicados pelo INEP (2019a), percebe-se que este cenário não é exclusivo da Unitins. Para o INEP (2019b) a evasão escolar e a permanência por longo tempo dos alunos nas instituições de ensino superior, tornaram-se uns dos grandes desafios para a educação brasileira. As buscas pelas causas têm se tornado objetos de pesquisas e estudos. De um lado são os investimentos do setor público em estruturas físicas, administrativas e acadêmicas, sendo pouco aproveitadas e sem o retorno devido, e do outro lado, estudantes que iniciam os estudos e não concluem (FILHO e colab., 2007).

As dificuldades de acesso e de se manter na escola é um cenário atual da educação brasileira. Segundo Censo da Educação Superior de 2018 dos alunos que ingressaram nas Universidades Públicas e Privadas no ano de 2010, 56,8% desistiram do curso e 37,9% conseguiram concluir o seu curso. Os demais que são os 5,3% que permaneceram na graduação por mais de 6 anos (INEP, 2019b). Outro fator importante e que precisa ser revertido é o tempo de estudo do brasileiro. A média no Brasil para a faixa etária entre 18 a 29 é de apenas 11,3 anos. Em se tratando da região Norte do país, este número é menor ainda, apenas 10,8 anos de estudos. Com estes números a região norte, ocupa a 4ª posição do ranking, ganhando apenas da região Nordeste que é de 10,7 (INEP, 2019a).

Especificamente o Estado do Tocantins, onde a Unitins está inserida, além da média de anos de estudos da região, observa-se que em 2018 ingressaram 830 alunos na Rede Federal de ensino, que vieram de outros estados para estudar no Tocantins (INEP, 2019a). O cenário aponta uma atenção especial, já que estudos do INEP evidenciam a desistência e trancamento para estes grupos de alunos (INEP, 2019a). A tabela 1 apresenta os dados da desistência e trancamento dos ingressantes que mudaram de Unidade Federativa em 2018.

**Tabela 1** -Ingressantes da rede federal que mudaram de Unidade da Federação (UF) em 2018

<b>Ano de Ingresso</b>	<b>Qtde de Alunos Ingressantes</b>	<b>Alunos matriculados em uma UF diferente de sua residência</b>	<b>Desistência</b>	<b>Trancamento</b>
------------------------	------------------------------------	--	--------------------	--------------------

2018	309.266	33.929	10,4%	4,1%
------	---------	--------	-------	------

Fonte: Elaborado pela autora com base em INEP (2019a)

Além disso, 18% dos alunos ingressantes em 2018 nas universidades federais, fizeram o Exame Nacional do Ensino Médio (ENEM) mais de uma vez, mesmo já frequentando alguma Instituição de Ensino Superior (IES). Alguns motivos podem ser justificados, como mudança de curso ou IES, fenômeno este que pode ser percebido, pela alta quantidade de vagas remanescentes e a evasão escolar (INEP, 2019a).

Os números indicam que a evasão escolar é um desafio que precisa ser superado e somente através de ações efetivas, este cenário poderá ser revertido. Já no que tange a legislação brasileira, cabe a família e o Estado a orientação a criança no seu percurso sócio-educacional, através da Lei 9.394/96 de Diretrizes e Bases da Educação Nacional, que é bem esclarecedora.

Art. 2º A educação, dever da família e do Estado, inspirada nos princípios de liberdade e nos ideais de solidariedade humana, tem por finalidade o pleno desenvolvimento do educando, seu preparo para o exercício da cidadania e sua qualificação para o trabalho (BRASIL, 1996c).

Outra ação governamental de grande relevância, foi a criação da Lei 10.287/2011 que altera o dispositivo da Lei 9.394/1996, instituindo competências para as instituições de ensino em notificar ao Conselho Tutelar do Município, juiz competente ou representante do Ministério Público, os alunos que estiverem com faltas acima de cinquenta por cento do percentual permitido em lei (BRASIL, 2001).

Percebe-se, que mesmo com as leis brasileiras para o combate desse fenômeno é necessário entender todo o contexto envolvido. Para MACHADO (2007) o sistema educacional é composto por partes inter-relacionadas e interativas. Os resultados positivos de uma política educacional podem ser concebidos por meio da construção entre os diversos níveis de ensino, particularmente entre a Educação Básica e o Ensino Superior.

Portanto, este estudo justifica-se pelo problema da evasão escolar que as IES estão enfrentando sobre esse fenômeno. A busca pelo diagnóstico mais preciso e premeditado, tornou-se algo desejado e estratégico. Além disso, espera-se que o resultado deste trabalho,

possa subsidiar a Unitins na identificação de padrões e modelos estatísticos dos perfis dos alunos evadidos, com informações capazes de influenciar em um processo decisório.

## **1.2 Objetivos**

### **1.2.1 Objetivo Geral**

O objetivo principal deste trabalho é pesquisar o problema de evasão escolar, propondo uma solução tecnológica envolvendo mineração de dados educacionais com aplicação de um estudo de caso na Universidade Estadual do Tocantins.

### **1.2.2 Objetivos Específicos**

- 1 Contextualizar o problema da evasão escolar, por meio de levantamento bibliográfico;
- 2 Identificar e discutir tecnologias registradas (patentes e registros de programas de computador) que abordam o problema abordado;
- 3 Propor uma solução de software que receba como entrada os dados acadêmicos e produza previsões a respeito da condição dos alunos em evadir, baseado em mineração de dados educacionais, tendo como base os dados acadêmicos da Unitins;
- 4 Realizar experimentos de validação da qualidade do preditor, à luz de métricas aplicadas a sistemas classificadores, tendo como base os dados acadêmicos da Unitins.

## **1.3 Estruturas do Trabalho**

O segundo capítulo apresenta o referencial teórico onde são descritos sobre o contexto da Evasão Escolar no Ensino Superior, com descritivos dos motivos e causas que levam alunos a desistirem do seu curso. Além disso, são apresentadas as ferramentas tecnológicas utilizadas, como a Mineração de Dados Acadêmicos e Aprendizado de Máquina e sobre o Sistema Inteligente.

O terceiro capítulo envolve toda parte metodológica, com a apresentação dos métodos e ferramentas utilizadas durante a pesquisa, bem como a amostra dos dados, a modelagem do sistema e as métricas utilizadas para a avaliação da Mineração de Dados.

No quarto capítulo, são realizadas buscas de patentes e programas de computador em bases de dados nacionais e internacionais com foco em tecnologias voltadas para a Evasão

Escolar. No quinto capítulo, apresenta-se a aplicação das técnicas de aprendizado de máquina, por meio de um estudo de caso dos dados acadêmicos da Unitins, permitindo assim a comprovação que essa técnica pode ser utilizada para identificar os alunos evadidos e não evadidos.

No sexto capítulo é o desenvolvimento de um sistema inteligente de diagnóstico preditivo dos alunos com risco de evasão e daqueles que estão propensos a graduar. Encontra-se neste capítulo a descrição da implementação deste sistema, que foi desenvolvido utilizando as técnicas de Aprendizado de Máquina.

E por fim o último capítulo é o sétimo, onde são apresentadas as considerações finais, as sugestões para trabalhos futuros e as produções bibliográficas geradas pela pesquisa.

## **2 REVISÃO DE LITERATURA**

Este capítulo aborda temas imprescindíveis que fundamentam esta pesquisa. Seus aspectos são apresentados, sem a intenção de esgotamento dos assuntos, que são voltados para a problemática da evasão escolar no ensino superior brasileiro. Nesta seção apresenta-se: A Evasão Escolar no ensino superior Brasileiro; Conceitos e aplicações da Inteligência Artificial; O uso da Mineração de Dados; A tendência da Mineração de Dados Educacionais; Conceitos e aplicações do Aprendizado de Máquina.

### **2.1 A Evasão Escolar no Ensino Superior Brasileiro**

A evasão escolar e a permanência por muito tempo dos estudos são barreiras para a Administração Pública que acarreta reflexos institucionais e sociais, por estarem associados a investimento em recursos humanos e financeiros (PEREIRA et al., 2011; BAGGI e colab., 2011). Estudos internacionais como de (SPADY, 1970; TINTO, 1975) já alertavam os pesquisadores quanto a necessidade maiores discussões e referenciais teóricos, para a elaboração de propostas que pudessem ser estudados por meios estatísticos. Já no Brasil os estudos sobre essa temática iniciaram a partir de 1995, organizado pelo Ministério da Educação (MEC) por parte de uma agenda governamental chamada de “Seminário sobre evasão nas universidades brasileiras”. Neste mesmo seminário, foi instituído a Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras (CEUPB) (KIPNIS, 2000).

A CEUPB foi criada em 1996, onde foi empreendido um dos primeiros esforços brasileiros para identificação das causas da evasão no País. Esta comissão teve como objetivo o estudo sobre os desempenhos das IES brasileiras, cuja formação era as Instituições de Ensino Superior Públicas, Federais e Estaduais (BRASIL, 1996b). Dentre os estudos desta comissão, foram definidas ações a serem desenvolvidas. O quadro 1, apresenta alguns objetivos desta comissão.

#### **Quadro 1 - Objetivos da CEUPB**

<b>Item</b>	<b>Ação</b>
-------------	-------------

1	Definir o conceito de evasão, considerando suas dimensões concretas: evasão de curso, evasão da instituição e evasão do sistema de ensino superior;
2	Definir e aplicar a metodologia igualitária de coleta e tratamento dos dados;
3	Identificar as taxas de diplomação, retenção e evasão dos cursos de graduação das IES do país;
4	Apontar as causas internas e externas da evasão, considerando as especificidades dos cursos e das regiões do país;
5	Definir estratégias voltadas à redução dos índices de evasão nas universidades públicas brasileira.

Fonte: Elaborado pela autora com base em (BRASIL, 1996b).

Neste sentido, o desdobramento da ação de número 1 resultou no quadro 2. O quadro 2 apresenta as definições sobre os tipos de evasão, permitindo assim, um melhor entendimento sobre o fenômeno. Segundo BRASIL, (1996b) a evasão é definida:

#### **Quadro 2 - Distinção sobre conceitos de Evasão Escolar**

<b>Tipo</b>	<b>Descrição</b>
Evasão do Curso	É quando o estudante é desligado do curso superior em situações diversas tais como: abandono (deixa de matricular-se), desistência (oficial), transferência ou reopção (mudança de curso), exclusão por norma institucional.
Evasão da Matrícula	É quando o estudante se desliga da instituição na qual está matriculado.
Evasão do Sistema	Quando o estudante abandona de forma definitiva ou temporária o ensino superior.

Fonte: Elaborado pela autora com base em (BRASIL, 1996b).

Além disso, a CEUPB em seu relatório intitulado “Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superiores Públicas” fez alguns apontamentos importantes e dentre eles é a falta de referenciais teóricos. Corroborando com

(SPADY, 1970; TINTO, 1975) a falta de referenciais teóricos limitou uma maior discussão a respeito do fenômeno. Além disso, outros pontos foram destacados como limitadores, as dificuldades operacionais como tempo, disponibilidades da comissão e dentre outros. O texto descrito no relatório da Comissão reforça essa necessidade:

O reconhecimento dos óbices que condicionaram este estudo corrobora a certeza de que o conhecimento mais completo e confiável do fenômeno só poderá ser alcançado através de um verdadeiro programa integrado de pesquisas que estabeleça os elos entre os níveis, identifique causas internas e externas, dando assim a necessária dimensão de totalidade característica de uma avaliação do sistema de ensino superior público do país. (BRASIL, 1996b)

Portanto, para uma melhor definição sobre os conceitos da evasão escolar, estudos mais atuais como de (BAGGI e colab., 2011) definem a evasão escolar como a saída do aluno da instituição antes da conclusão do seu curso. Percebe-se que esta definição corrobora com a Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras (BRASIL, 1996b). Por outro lado (ABBAD e colab., 2006) define a evasão como a desistência definitiva do aluno em qualquer etapa do curso. E para um melhor entendimento, verifica-se a presença de estudos mais recentes como de (FILHO e colab., 2007); (RIGO e colab., 2014);(COLPANI, 2019) que buscam sobretudo o entendimento sobre a evasão, seja ela através de conceitos e/ou indicadores.

Para (FILHO e colab., 2007) a análise da evasão estudantil tem que partir do ponto de vista institucional, onde os resultados afetam os sistemas educacionais, causando ociosidade de professores, funcionários, equipamentos e espaços físicos. Segundo os autores este problema de ociosidade está presente tanto nas IES públicas e privadas. (RIGO e colab., 2014) os estudos sobre a evasão escolar tem possibilitado a identificação da relação das demandas junto a sociedade, fenômeno este, que está presente tanto nas universidades públicas e privadas e a sua consequência são as questões financeiras e a redução de número de alunos formados no ensino superior, causando impacto na cadeia produtiva do nosso país.

COLPANI, (2019) utilizou em seus estudos os dados do CENSO e identificou que em 2017 o Estado do Pará foi o maior em índices de evasão escolar no Brasil. Em sua pesquisa, foi diagnosticado que a média de evasão nas escolas do Estado é de 20%. Deste percentual, estão um uma escala de 38% a 65% dos alunos com idade acima da recomendada que é de 18



anos, previstas na Lei 9.394/96 (BRASIL, 1996c). Assim, o indicador da taxa de distorção de série e idade, foi a variável que mais se associou com a evasão escolar.

Assim, considera-se que essa temática ainda requer estudos mais aprofundados, uma vez que conceitos e metodologias ainda se divergem. Necessário o desenvolvimento de estudos que busquem melhores entendimentos para uma discussão mais sólida. A realização de estudos que busquem responder a estas indagações é essencial para o estabelecimento de um campo de discussão, auxiliando as IES nas políticas educacionais, contribuindo assim, para uma educação superior de qualidade no para o nosso país.

## **2.2 Conceitos e a aplicações da Inteligência Artificial**

Segundo (PITEIRA e colab., 2019) a Inteligência Artificial (IA) está surgindo com o um novo mundo imaginário e a cada dia está ganhando papel de destaque para as organizações, principalmente para a tomada de decisão. Com uma gama de informações e conhecimentos, onde está dinâmica influenciará a tomada de decisão dos mais variados setores, como bancário, automóvel e na educação. Este novo mundo, então criado por algoritmos, executam tarefas sociais e cognitivas, que antes eram realizadas por humanos. Para (DE MEDEIROS, 2018) de todas as tecnológicas, técnicas e ferramentas que foram desenvolvidas ao longo da história, a Inteligência Artificial (IA) foi uma das conquistas mais representativas para humanidade.

De acordo com SOWA (1984) a IA não é algo novo, ela estava presente desde a antiguidade, por meio das teorias de raciocínio e aprendizado, nas perspectivas de como a mente é constituída pela operação de um sistema físico. Para o autor a IA se remete a tradicional escola Filosófica, através de dois grandes precursores, Aristóteles por meio das aplicações, refinamento e reinvenção e Leibniz pelo desenvolvimento da quantificação do conhecimento e raciocínio do trabalho, chamado *Characteristica Universalis*, quando fez o seu primeiro proponente da IA. Na definição de SOWA (1984) a IA foi baseada em dois campos, o primeiro é a da ciência cognitiva, por meio da filosofia, linguística e da psicologia e com forte influência da ciência da computação e o segundo campo, é das engenharias que servem como um complemento para IA.

Mas foi no ano de 1956, que o termo IA passou-se a ser utilizada pelo cientista John McCarthy, em um projeto de pesquisa liderado por ele e mais três cientistas Marvin Minsky, Nathaniel Rochester e Claude Shannon na Universidade de *Dartmouth* nos Estados Unidos. O projeto original não era visto com uma máquina inteligente, mas a Inteligência com vários recursos, com o uso do computador para simular cada um deles. Além disso, as propostas das técnicas não estavam restritas apenas a computação simbólica, já que McCarthy era um matemático, e sim para os diferentes aspectos da Inteligência Artificial (GANASCIA, 2010). Nesta mesma conferência de verão de *Dartmouth*, foram elencados alguns temas e aspectos que poderiam ser considerados problemáticos para a IA. O quadro 3 apresenta os 7 aspectos abordados na conferência.

**Quadro 3** - Temas do Congresso em 1956 sobre a Inteligência Artificial

Item	Aspecto	Descrição
1	Computadores automáticos	Se uma máquina pode fazer um trabalho, em seguida, uma calculadora automática pode ser programada para simular a máquina. Onde a velocidade e capacidade de memória dos computadores atuais podem ser insuficientes para simular muitas das funções superiores do cérebro humano, mas o maior obstáculo não é a falta de capacidade da máquina, mas a nossa incapacidade de escrever programas aproveitando ao máximo o que temos.
2	Como um computador pode ser programado para usar uma linguagem	Especula-se que grande parte do pensamento humano manipula as palavras de acordo com as regras da razão e da especulação. A partir deste ponto de vista, forma-se uma generalização e admite-se uma nova palavra e algumas regras por meio de sentenças. Essa ideia nunca havia sido formulada ou nem exemplos tirados do papel.

3	Redes Neurais	Um conjunto hipotético de neurônios, pode ser arranjada de forma que, reunida, crie um conceito. É em cima deste conceito, criado pelo grupo de Uttley, Rashevsky, Farley, Clark, Pitts, McCulloch, Minsky, Rochester e Holland têm se debruçado, produzindo uma quantia considerável de trabalho teórico e experimental. Resultados parciais têm sido obtidos, mas o problema precisa de mais trabalho teórico.
4	Teoria do tamanho do cálculo	Se fosse dado um problema definido se for possível testar mecanicamente se a resposta proposta é válida ou não, uma maneira de resolver é tentar todas as respostas possíveis na ordem. Esse método não é eficaz. E para eliminá-lo é necessário ter alguns critérios para a eficiência do cálculo.
5	Auto aperfeiçoamento	Uma máquina inteligente executará atividades por conta, o que provavelmente pode ser descrito como auto aperfeiçoamento. Percebemos que vale a pena realizar estudos mais aprofundados sobre esse tema.
6	Abstrações	Uma quantidade de abstrações pode ser distintamente definida, enquanto outras não. Uma tentativa é a classificação e descrição dos métodos das máquinas na formação dessas abstrações, a partir de sensores e outros dados que valem a pena o aprofundamento.
7	Aleatoriedade e criatividade	Uma conjuntura bastante atrativa e ainda bem incompleta é a diferença entre pensamento criativo, pensamento competente e o inimaginável e que se mantém sobre a ideia de aleatoriedade. A aleatoriedade deve ser guiada pela intuição para ser eficiente. Em outras palavras, o palpite inclui aleatoriedade controlada em vez de pensamento ordenado.

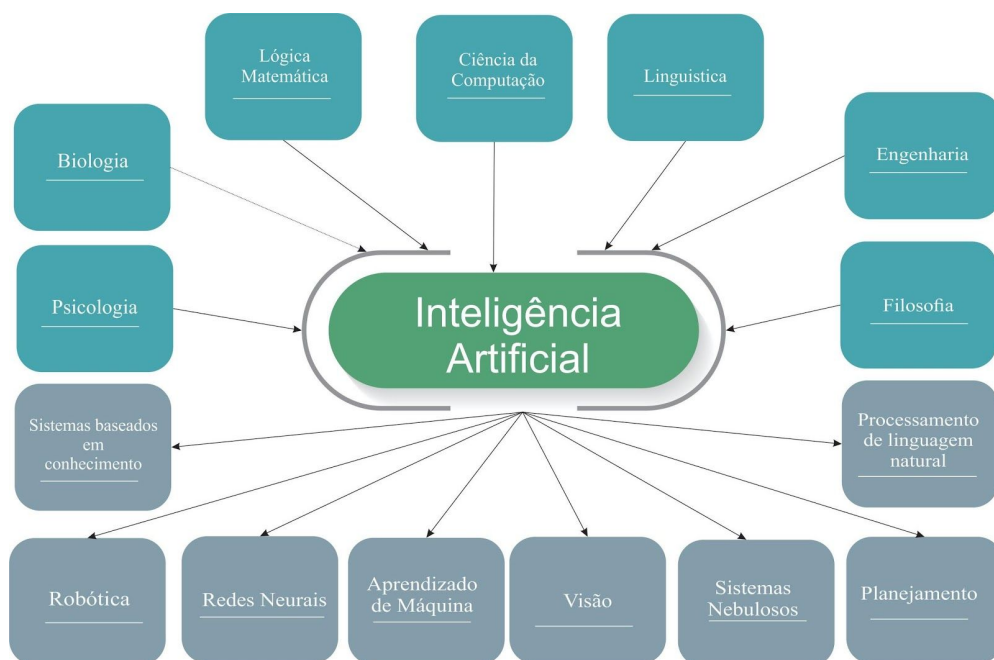
Fonte: Elaborado pela autora com base em (MCCARTHY e colab., 2006)

De acordo com as revisões das literaturas, percebe-se que a IA além de fazer parte do ramo da ciência da computação, ela está presente também nos campos da psicologia, biologia,

lógica matemática, linguística, engenharia e filosofia. A figura 1, apresenta as áreas que se relacionam.

**Figura 1** - Áreas de relacionamento com a Inteligência Artificial

### Áreas da Inteligência Artificial



Fonte: Elaborado pela autora com base em (MONARD, BARANAUKAS, 2000)

Em se tratando de definições e conceitos, segundo (DE MEDEIROS, 2018) não há um consenso único para a definição da IA, pois existem diversos elementos que se manifestam de formas variadas em razão de interpretações distintas sobre os processos de IA, que se relacionam com os mecanismos do cérebro e da mente humana. Por sua vez, MCCARTHY, (1956) define que é uma ciência que estuda o comportamento da inteligência humana, por meio de máquinas (GANASCIA, 2010). De acordo com (DE MEDEIROS, 2018) a inteligência artificial é definida através do ramo da Ciência da Computação, que se ocupa da computação inteligente.

Contudo, sob o ponto de vista dos autores (RUSSELL e colab., 2010) são apresentados alguns conceitos ao longo da história da IA e que estão divididas em 4

categorias. As abordagens das categorias 1 e 3 estão relacionadas ao processo de pensamento e raciocínio, enquanto as 2 e 4 são referentes ao comportamento. O quadro 4 demonstra as abordagens sobre os conceitos de IA.

**Quadro 4** - Abordagens sobre o conceito da Inteligência Artificial

Item	Categoria	Conceito
1	Sistemas que pensam como seres humanos	O novo e interessante esforço para fazer os computadores pensarem... máquinas com mentes, no sentido total e literal". (HAUGELAND, 1985)
2	Sistemas que atuam como seres humanos	A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas. (KURZWEIL, 1990)
3	Sistemas que pensam racionalmente	O estudo das faculdades mentais pelo seu uso de modelos computacionais (CHARNIAK; MCDERMOTT, 1985)
4	Sistemas que atuam racionalmente	Inteligência Computacional é o estudo do projeto de agentes inteligentes (POOLE et al., 1998)

Fonte: Elaborado pela autora com base em (RUSSELL e colab., 2010)

Neste sentido, é perceptível que as tecnologias estão transformando a sociedade, com mudanças significativas no nosso dia a dia, que vai desde a forma de se comunicarmos até as funções exercidas pelas empresas, com processos mais ágeis e eficazes. Portanto, a Inteligência artificial permite que as tarefas mais complexas sejam viabilizadas por máquinas de forma que as máquinas pensem como seres humanos, através da capacidade de aprender, raciocinar de decidir logicamente sobre algo, para as soluções de problemas.

### 2.3 O uso da Mineração de dados

Com os avanços das Tecnologias de Informação (TI), trouxeram novas concepções para a sociedade. De acordo com (KOHN e MORAES, 2007) as riquezas dos países passaram a ser o acesso às tecnologias e a sua capacidade de desenvolvimento. Onde a informação é transformada em conhecimentos e estratégias de negócios para um objetivo único. Mas esse

avanço trouxe algumas consequências, como o aumento expressivo no volume dos dados, gerando assim, a necessidade do armazenamento de grandes massas de dados. Tecnologias como internet, memórias secundárias com maiores capacidades e sistemas de informações gerenciais, fez com que se proliferasse inúmeras bases de dados (GOLDSCHMIDT e PASSOS, 2005). A partir destes “sintomas”, surgem as ferramentas computacionais automáticas e inteligentes, que auxiliam as organizações nas tarefas de analisar, interpretar e relacionar dados para a geração do conhecimento. Conceitos estes trabalhados através das técnicas da Mineração de Dados, que visa a exploração dos dados a fim extrair padrões consistentes dos dados.

Para HIRJI (2001) a Mineração de Dados (MD) pode ser definida como uma análise não trivial de dados, por meio dos bancos de dados, com a finalidade da descoberta de informações novas. Essa descoberta é feita por meio de padrões e regras a partir de relacionamentos entre os elementos dos dados. Segundo HIRJI (2001) a MD surgiu com uma classe das técnicas analíticas, que vão além da estatística, visando a análise de grande quantidade de dados. Na definição de (CABENA, 1998) a MD está relacionada aos conhecimentos da análise estatística de dados, com aprendizagem de máquina, reconhecimento de padrões e visualização dos dados. De acordo com (COSTA e colab., 2013) a Mineração de dados pode ser vista como uma das etapas do processo da Descoberta de Conhecimento em Banco de Dados ou *Knowledge Discovery in Databases* (KDD), cujo objetivo é a descoberta de conhecimento e que possa ser útil, que traga algum benefício para uma possível tomada de decisão.

No entanto, nas abordagens dos autores sobre a definição de MD é perceptível a relação entre a estatística por meio das relações entre os dados. Já a IA através das máquinas que assemelham os humanos e o aprendizado de máquina com algoritmos que aprendem com os dados, para que sejam feitas as previsões. Além disso, podemos encontrar estudos científicos, como de (RAMESH e colab., 2013; KAMPFF et al., 2014; BRITO et al., 2015) que utilizaram estas técnicas como forma de melhoria para a qualidade do ensino.

Os autores (RAMESH e colab., 2013) ressaltam sobre a necessidade e a importância das técnicas de mineração de dados, como mecanismos para auxiliar a qualidade dos dados dos sistemas acadêmicos. Utilizados com o foco de identificação de riscos que possam ser

minimizados para a melhoria da qualidade do ensino. Nos estudos como de (KAMPFF et al., 2014) utilizaram a Mineração de Dados em um contexto de EAD, como forma de descobrir padrões de evasão e reprovação. A MD permitiu que fossem gerados alertas, para os professores que tinham alunos nos grupos de risco. Para a descoberta do conhecimento foram utilizados 87 atributos, como: sexo, idade, nota de cada avaliação da disciplina, número de acessos ao ambiente virtual, conteúdo online, biblioteca virtual, fórum e dentro outros. Como resultado, foi identificado que percentual de reprovações, para aqueles alunos que receberam as alertas por parte dos professores, foram menor do que não receberam.

(BRITO et al., 2015) utilizou as técnicas de MD no curso de bacharelado em Ciência da Computação da Universidade Federal da Paraíba (UFPB). A escolha do curso se deu devido a alta taxa de evasão dos cursos de computação. Os atributos escolhidos, foram as notas do vestibular e das disciplinas do primeiro período. O modelo foi baseado nos dados históricos dos alunos entre os anos de 2001 até 2013. O estudo também apresentou a definição da UFPB sobre os requisitos que consideram um aluno evadido. Portanto, um aluno é evadido nas condições em que o mesmo não efetivou a sua matrícula nas disciplinas, nem realizou o trancamento total do curso ou não conseguiu concluir o seu curso no prazo máximo de 6 anos. O resultado deste estudo foi com o uso do método de aprendizado de máquina, onde as taxas geradas chegaram até 86,9%. Resultado este, que indica uma abordagem viável para a detecção de grupos de risco.

Para (HIRJI, 2001) a Mineração de Dados “não é o fim, mas o meio para um fim”, através de descobertas de dados estratégicas para um objetivo. Neste sentido, a frase de (HIRJI, 2001) faz com que tenhamos a percepção que o uso das técnicas de Mineração de Dados, vai muito além da sua operacionalização, onde seus resultados são grandes fontes estratégicas para a tomada da decisão.

## **2.4 A Tendência da Mineração de dados Educacionais**

De acordo com ([RIGO e colab., 2014](#)), o aumento substancial das tecnologias de informação e comunicação, possibilitou o aumento das bases de dados e a sua disponibilização, fazendo com que a capacidade de geração dos dados, fosse maior do que a capacidade dos pesquisadores e analistas. Essas limitações podem ser ocasionadas por diversos fatores, como limitações técnicas, dados não estruturados e dentro outros aspectos.

Este cenário de grande disponibilidade do conjunto de dados está presente também na área de educação, onde as instituições de ensino possuem uma gama de sistemas informatizados, como ambiente de educação a distância, sistemas acadêmicos e comunidades virtuais. A partir deste cenário, a busca pela exploração dos dados dos ambientes educacionais por novas descobertas, fez surgir uma nova área de pesquisa chamada de Mineração de Dados Acadêmicos (MDE) ou *Educational Data Mining*, (EDM).

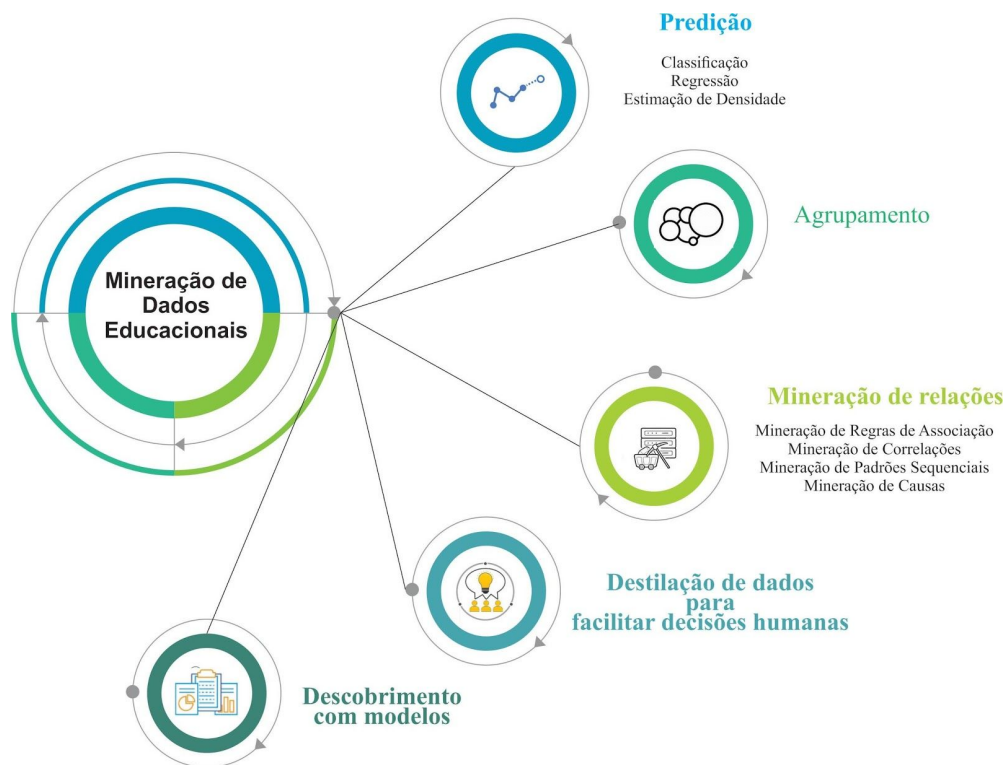
A Mineração de Dados Educacionais, nada mais é do que a aplicação das técnicas Mineração de dados, em um conjunto de dados, para os diversos contextos dos sistemas educacionais Sendo que a natureza destes dados é mais variada do que a observada nos dados tradicionais, com a necessidade de ajustes e novas técnicas. Os autores como (RIGO e colab., 2014) destacam que esta diversidade de dados, agregam um grande potencial de recursos, para o auxílio na melhoria da educação . Para (ROMERO e colab., 2010) a MDE é o processo de conversão dos dados educacionais em informações, para a tomada de decisão. Onde a técnica visa a mineração dos dados nos diversos sistemas informatizados, em busca de padrões significativos a fim de tornar a informação útil tanto para os alunos e professores.

Sob o ponto de vista de (BAKER e colab., 2011) até o ano de 2011 este tipo de técnica no campo educacional ainda era pouco explorada no Brasil, países como Estados Unidos, Canadá e Espanha, estavam na frente com as pesquisas relacionadas a este tema. Nos estudos realizados pelos autores (BAKER e colab., 2011) é possível identificar as principais tarefas que melhores se adaptam a MDE. Os autores ressaltam que cada técnica tem a sua peculiaridade e fica a cargo do cientista de dados a escolha das técnicas a serem utilizadas na sua análise. A figura 2 apresenta a taxonomia proposta pelos autores das subáreas da pesquisa em MDE.

**Figura 2-** Taxonomia das principais subáreas da pesquisa em MDE

### **Taxonomia da Mineração de Dados Educacionais**





Fonte: Elaborado pela autora com base em (BAKER e colab., 2011)

Percebe-se, o crescente esforço dos pesquisadores que buscam o conhecimento das técnicas de MDE como forma de auxiliar o ensino superior brasileiro. Os autores como (RIGO e colab., 2014) enfatizam que o uso das técnicas de MDE mitigam a evasão escolar, por meio de disponibilização de informações a respeito dos alunos para fins pedagógicos. Segundo (BAKER e colab., 2011) os autores entendem que a MDE tem muito a contribuir para educação brasileira com a geração do conhecimento em diferentes domínios, tanto no ensino quanto na aprendizagem. Em se tratando do crescimento do ensino a distância (EAD) a MDE também pode proporcionar impactos positivos, com a análise dos dados e na criação de modelos que compreendam os processos de aprendizagem.

Corroborando com (BAKER e colab., 2011) nos estudos realizados por (DE SOUZA BRITO e colab., 2020) buscou-se a identificação da evasão escolar, por meio da Análise de Aprendizagem ou termo em inglês *Learning Analytics* dos riscos de evasão dos alunos do Instituto Federal da Paraíba (IFPB). Os cursos escolhidos foram os técnicos profissionalizantes e superiores nas áreas de Informática e Computação, superiores de Licenciatura em Letras e Administração. Foi desenvolvido um programa de computador

*plugin* (módulo de extensão) do tipo *report* (relatório) no Ambiente Virtual de Aprendizagem *Moodle*, que monitorava a base de dados de alunos, interações nos fóruns, acessos, notas das atividades. O monitoramento se relacionava aos indicadores cognitivos, sociais e comportamentais dos estudantes. Como resultado, este *plugin* auxiliou a gestão acadêmica no monitoramento dos estudantes que estavam no grupo de risco de evasão. Os relatórios gerenciais forneceram os acessos dos alunos, desempenho e interações, melhorando assim, a tomada de decisão.

## **2.5 Conceitos e aplicações do Aprendizado de Máquina**

O Aprendizado de Máquina (AM) é a construção de um sistema, capaz de adquirir conhecimentos de forma automática (MITCHEL, 1997). (MONARD e BARANAUSKAS, 2003) e a sua construção é baseada em experiências acumuladas através da solução de problemas anteriores. Segundo (MONARD e BARANAUSKAS, 2003) o AM são técnicas computacionais, construídas por meio de desenvolvimento de sistemas de computador, que trabalham de forma automática. Aprendem a tomar decisões baseadas em experiências acumuladas de uma solução bem-sucedida de problemas anteriores. Cada sistema de aprendizagem tem a sua especificidade, não havendo um algoritmo único, capaz de solucionar todos os tipos de problemas. É necessário compreender as limitações e definir uma melhor metodologia, que permita avaliar os algoritmos para determinados tipos de problemas.

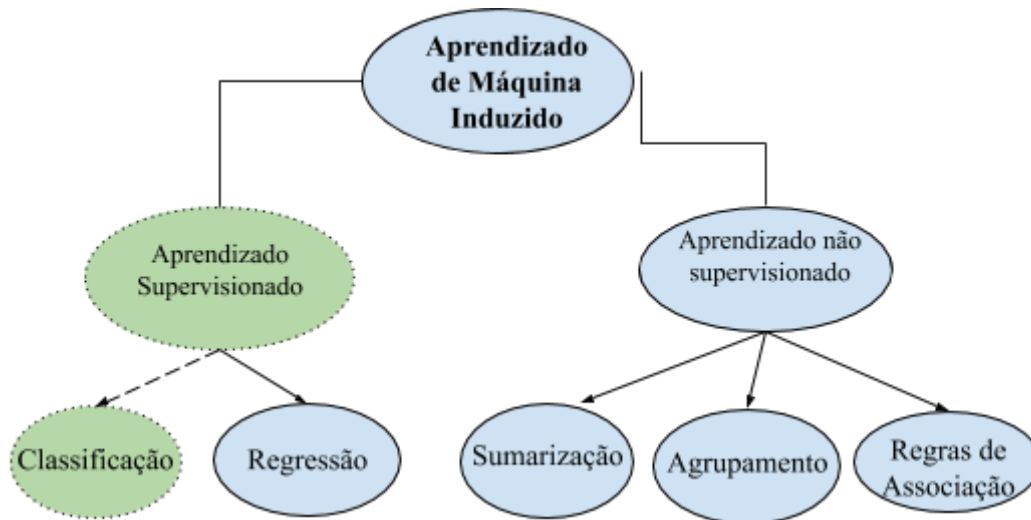
Para os autores (MONARD e BARANAUSKAS, 2003) na hierarquia do aprendizado, geralmente é utilizada a Indução para avaliar um algoritmo de aprendizado. A indução nada mais é que a obtenção de conclusões genéricas a partir de um conjunto particular de exemplos. Ela é determinada como o raciocínio que se origina de um conceito específico e o generaliza, isto é, da parte para o todo. Assim sendo, a hipótese fornecida por meio da inferência indutiva pode ser verdadeira ou não. Mesmo nesta condição, a inferência indutiva é um dos principais métodos conhecidos para advir o conhecimento novo ou predizer eventos futuros.

Sobretudo, o sistema de aprendizado indutivo ele é dividido em supervisionado e não supervisionado, como ilustrado na figura 3. A escolha supervisionado ou não supervisionado dependerá das tarefas e exemplos estarem ou não rotulados com o atributo de classe. Para

esta pesquisa, utilizaram-se as tarefas do aprendizado supervisionado, especificamente as que estão destaque.

**Figura 3** - Hierarquia do Aprendizado de Máquina

### A Hierarquia do Aprendizado de Máquina



Fonte: Elaborado pela autora com base em (MARTINS, 2003)

Os autores (MONARD e BARANAUSKAS, 2003) definem que o aprendizado supervisionado é fornecido ao algoritmo de aprendizado, ou indutor, um conjunto de exemplos de treinamento para os quais os rótulos da classe associada são conhecidos. Se as classes possuírem valores discretos o problema é denominado como classificação. Quando os valores são contínuos o problema é reconhecido como regressão. Ao contrário dos métodos supervisionados, como algoritmos de classificação, não há classes ou rótulos predefinidos para treinamento de um modelo, ou seja, o aprendizado é realizado de forma não supervisionada (HAN e colab., 2011). Assim, as tarefas podem estar relacionadas com o agrupamento dos exemplos, descrição compacta de um subconjunto de dados, no qual é chamado de sumarização ou com a caracterização. As regras de associação do quanto a presença de um agrupamento de atributos, acarreta na presença de algum outro conjunto específico de atributos do mesmo exemplo (MARTINS, 2003).

Estudos atuais como de (BARROS e colab., 2020) utilizaram os mesmos métodos adotados nesta pesquisa. De acordo com os autores a MDE utiliza de técnicas de aprendizado

de máquina, para soluções de problemas que envolvem estatísticas. A pesquisa realizada por (BARROS e colab., 2020) entre os anos de 2017 e 2018, identificaram os fatores de reprovação dos alunos do curso de Bacharelado em Ciências e Tecnologia da Universidade Federal do Rio Grande do Norte (UFRN). As disciplinas investigadas do 1º período foram: Pré-Cálculo, Cálculo I, Química Geral, Práticas de Leitura e Escrita I, Ciência, Tecnologia e Sociedade e Vetores e Geometria Analítica. E apenas a disciplina lógica de programação do 2º período. Os algoritmos utilizados foram IBk, *RandomForest* (RF), *BayesNET* (BNet) e *MultilayerPerceptron* (MLP). Os resultados apontaram taxas de acerto superiores a 80%, apontando os estudantes com tendência a evasão. As informações utilizadas como desempenho foi a prova de ingresso e nas disciplinas pré-requisito da disciplina de programação. O algoritmo acertou em 74,7% de acertos, tendo a maior média de precisão.

### 3 METODOLOGIA

Conforme elucidado na Introdução, esta pesquisa buscou através de ferramentas computacionais inteligentes, o estudo de métodos para a identificação de padrões que possam auxiliar os tomadores de decisão na identificação precoce dos alunos com risco de evasão. Alguns percursos metodológicos foram necessários para uma melhor compreensão e entendimento para este fenômeno, bem como os dados acadêmicos da Unitins, trabalhos acadêmicos e até mesmo soluções computacionais como patentes e registros de softwares, que também auxiliaram no desenvolvimento das metodologias. Para obtenção do acesso aos dados da Unitins, esta pesquisa foi submetida e aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos (CEP), sob o número 29598820.3.0000.8023.

#### 3.1 Métodos e Ferramentas

Para um melhor entendimento sobre a evasão escolar e dada a relevância sobre o problema, a pesquisa adotou o caráter exploratório, com a natureza quantitativa pela própria natureza dos dados. O uso das tecnologias de inteligência artificial teve como objetivo a identificação de regras passíveis de testes e aplicações em uma análise dos dados acadêmicos da Unitins. Tendo como proposta o desenvolvimento de metodologias baseadas em mineração de dados acadêmica (MD) com métodos de aprendizado de máquina (AM).

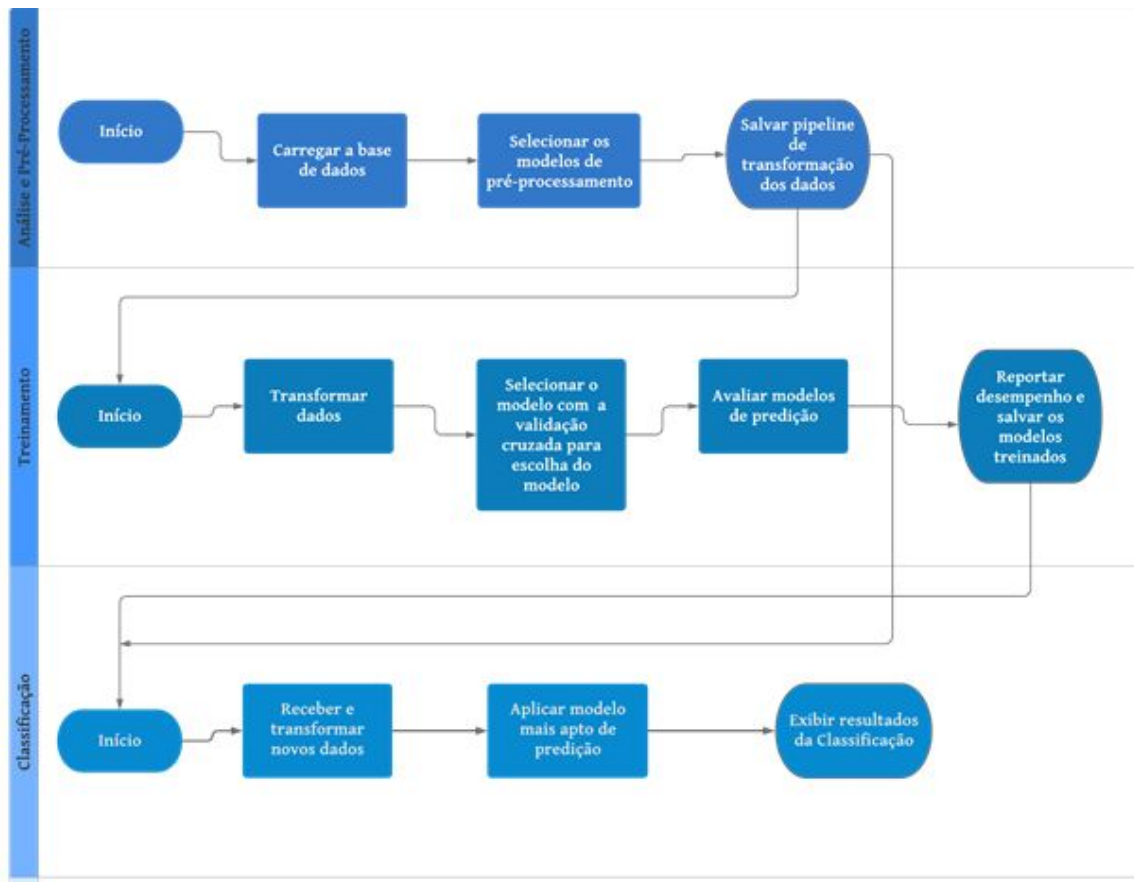
Os processos executados foram de coleta, análise e interpretação dos dados. Segundo CRESWELL (2010) estes processos têm o intuito de descobrir relações entre as variáveis e características de causalidade sobre ao fenômeno da evasão escolar. Como fundamentação teórica e técnica, foram realizados os levantamentos bibliográficos que fundamentam as metodologias propostas, bem como o uso de materiais elaborados e previamente publicados em livros e artigos científicos (GIL, 2008).

As ferramentas técnicas utilizadas para a aplicação das metodologias foram: *Spyder*, a linguagem de programação *Python* e os *frameworks Scikit-learn* e *Pandas*. Nota-se que trabalhos como de (JADRIC et al. (2010); SOUZA, 2008) utilizam-se de técnicas de mineração de dados, regressão logística e árvores de decisão, para diagnosticar a evasão escolar no ensino superior.

### 3.2 A modelagem para a escolha dos melhores atributos

A solução proposta para esta pesquisa avaliou os modelos de predição de aprendizado supervisionado. Para este modelo escolhido são construídos modelos para prever um atributo não conhecido de acordo com os atributos observados. Sendo assim, este trabalho apresenta um método de seleção que está apresentado na figura 4. Diferentemente do supervisionado, os métodos de aprendizado não supervisionados extraem padrões, como agrupamentos, gráficos de processo e correspondências entre os dados (DUAN e DA XU, 2012). Neste sentido, buscou-se a avaliação de preditores tais como Árvores de Decisão, Florestas Aleatórias, Redes Neurais Artificiais, Máquinas de Vetores de Suporte e Regressão Logística. Autores como (TURBAN e colab., 2009) classificam de forma ampla os algoritmos de Mineração de Dados em: classificação, agrupamento, associação e descoberta de sequência. A seguir, estarão sendo apresentados os processos executados na modelagem dos atributos.

Figura 4 – Modelagem dos Atributos



Fonte: Elaborado pela autora (2019)

O fluxograma apresentado demonstra as etapas gerais de um sistema de aprendizado de máquina. A primeira etapa, **Análise e Pré-processamento**, consistiu em efetuar processos de limpeza e adaptação da base de dados fornecida pelo sistema educacional da Unitins. Os dados originais estavam distribuídos em 3 tabelas, sendo: Cadastro de Alunos, Matrizes Curriculares e Histórico Acadêmico. O quadro 5 representa os campos da tabela de cadastro de alunos.

**Quadro 5** - Descrição dos campos da tabela de Cadastro de Alunos

<b>Campo Coluna</b>	<b>Descrição</b>
RA	Número do registro acadêmico
DTNASCIMENTO	Data de nascimento
IDADE	Idade
SEXO	Sexo
ESTADONATAL	Estado de nascimento
NATURALIDADE	Cidade de nascimento
NACIONALIDADE	Nacionalidade
ESTADOCIVIL	Estado civil
CORRACA	Cor
ESTADOENDERECO	Estado
CIDADEENDERECO	Cidade
TIPOINST2GRAU	Tipo de instituição que cursou no 2ª grau (pública/privada)
ANOULTIMAINST	Último ano cursado no ensino médio
CODCURSO	Código do curso
SEMESTRE_INGRESSO	Semestre de ingresso na IES

TIPOINGRESSO	Forma de ingresso (vestibular, SISU, transferência interna, externa, ex-officio e mandado de segurança e dentre outros)
CODGRADE	Código da matriz curricular
CODHABILITACAO	Código da habilitação do aluno ao curso
STATUS_ATUAL	Status atual da matrícula no curso (matriculado, trancado, desistente e dentre outros)
DTCOLACAOGRAU	Data da colação de grau
ULTIMO_SEMESTRE	Último semestre cursado
CAMPUS	Nome do câmpus
CURSO	Nome do curso
TURNO	Turno do curso
IDHABILITACAOFILIAL	Identificação da habilitação do aluno ao curso

Fonte: Elaborada pela autora (2020)

Na tabela de Cadastro de Alunos havia 5018 instâncias, que estão apresentadas por meio da figura 5. Para a etapa do pré-processamento não houve a exclusão de nenhum registro por motivo de valor nulo. Mas ao aplicar o filtro no campo STATUS\_ATUAL, para os tipos de status (matriculado, graduado, desvinculado e desistente) que foram considerados mais relevantes para a pesquisa, esse número foi reduzido para 4298 alunos. Ou seja, existiam 1979 alunos com status matriculados, 1279 alunos com status graduado e 1040 alunos com status desvinculado e desistente. Ao analisar os resultados houve a necessidade de uma nova remoção, a retirada do status matriculado, já que os mesmos não poderiam participar do treinamento, por não possuir um status final de formado ou evadido. Com a retirada deste status, resultou em um total de 2319 que fizeram parte dos treinamentos. Por fim, foi criado um novo status, denominado Evadido, a partir da junção dos status Desistente e Desvinculado e o status graduado passou a se chamar Formado, para não gerar quaisquer ambiguidades com o status dos dados originais.

**Figura 5** - Quantitativos de alunos por *Status Acadêmico*





Fonte: Elaborado pela autora com base em UNITINS (2020b)

Em seguida, nesta mesma etapa foi realizado o preenchimento dos valores para os atributos que não estavam preenchidos. Para o atributo idade foi definida a média de idade da base de dados, os demais atributos foram registrados como “não informado”. As próximas tabelas a serem pré-processadas foram de Matriz Curricular e Histórico acadêmico. Os quadros 6 e 7 apresentam essas tabelas.

O quadro 6 é a tabela Matriz Curricular que apresenta os campos referentes aos componentes curriculares dos cursos. Nessa tabela encontravam-se 2.064 instâncias. O quadro 7 apresenta a última tabela com a relação dos campos de Histórico Acadêmico. O total de instância desta tabela era de 122.737, referente aos dados de notas, frequências, semestre e dentre outros.

**Quadro 6-** Descrição dos dados da tabela Matriz Curricular

<b>Campo Coluna</b>	<b>Descrição</b>
CURSO	Nome do curso
DISCIPLINA	Nome da disciplina
CODGRADE	Código da matriz curricular
CAMPUS	Nome do câmpus

CODHABILITACAO	Código da habilitação do curso
CODDISC	Código da disciplina
CODPERIODO	Período da disciplina na matriz curricular
CARGAHORARIADISCIPLINA	Carga horária da disciplina
CARGAHORARIA	Carga horária total da matriz curricular
TURNO	Turno do curso na matriz curricular
IDHABILITACAOFILIAL	Identificação da habilitação do curso

Fonte: Elaborada pela autora (2020)

**Quadro 7-** Descrição dos dados da tabela de Histórico Acadêmico

<b>Campo Coluna</b>	<b>Descrição</b>
RA	Número do registro acadêmico
CODCURSO	Código do curso
CAMPUS	Nome do campus
CURSO	Nome do curso
DISCIPLINA	Nome da disciplina
STATUSFINAL	Status final (aprovado, reprovado por média, reprovado por falta, desistente e dentre outros)
NOTAFINAL	Média final
FALTAFINAL	Quantidade total de faltas
NOTAFALTA	Média final
NOTAFALTA2	Quantidade total de faltas

IDTURMADISC	Identificação da disciplina ofertada no semestre
CODDISC	Código da disciplina
CODTURMA	Código da turma
SEMESTRE	Ano e semestre que a disciplina foi cursada

Fonte: Elaborada pela autora (2020)

O próximo passo foi a extração dos dados não implícitos que foram extraídos a partir do histórico acadêmico e da matriz curricular, na busca da identificação dos totalizadores, como: disciplinas da matriz, disciplinas cursadas, disciplinas aprovadas e disciplinas reprovadas por médias e faltas. De acordo com (KOHAVI; JOHN, 1997) a técnica de criar e de selecionar atributos, permite que sejam selecionados dados mais eficazes e também a redução de dimensionalidade dos dados, facilitando a execução do algoritmo de mineração de dados. O quadro 8 apresenta o resumo das etapas seguidas pela pesquisa no pré-processamento.

#### **Quadro 8-** Descrição dos tipos de pré-processamento

<b>Campo</b>	<b>Tipo de pré-processamento</b>
IDADE	Substituição dos valores nulos pela média da idade e escalonamento dos valores.
SEXO	Substituição dos valores nulos pelo valor 'I', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
ESTADOCIVIL	Substituição dos valores nulos pelo valor 'I', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
CORRACA	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
ESTADADONATAL	Substituição dos valores nulos pelo valor 'NI', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
NATURALIDADE	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.

ESTADOENDERECO	Substituição dos valores nulos pelo valor 'NI', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
CIDADEENDERECO	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
TIPOINST2GRAU	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
ANOULTIMAINST	Substituição dos valores nulos pelo valor '0000', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
SEMESTRE_INGRESSO	Substituição dos valores nulos pelo valor '00000', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
TIPOINGRESSO	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
CODGRADE	Substituição dos valores nulos pelo valor '0' e escalonamento dos valores.
ULTIMO_SEMESTRE	Substituição dos valores nulos pelo valor '00000', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
CAMPUS	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
CURSO	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
TURNOS	Substituição dos valores nulos pelo valor 'NÃO INFORMADO', Codificação de rótulo para os valores numéricos e escalonamento dos valores.
total_disciplinas	Campo criado tomando-se a soma das disciplinas da tabela Matriz Curricular referente ao curso do aluno e escalonamento dos valores.
disciplinas_cursadas	Campo criado tomando-se a soma de disciplinas da Tabela Histórico já vencidas pelo aluno e escalonamento dos valores.
disciplinas_aprovadas	Campo criado tomando-se a soma das disciplinas da Tabela Histórico Acadêmico que o aluno foi aprovado e escalonamento dos valores.
disciplinas_reprovadas_nota	Campo criado tomando-se a soma das disciplinas da Tabela Histórico Acadêmico que o aluno foi reprovado por nota e escalonamento dos valores.

disciplinas_reprovadas_falta	Campo criado tomando-se a soma das disciplinas da Tabela Histórico Acadêmico que o aluno foi reprovado por falta e escalonamento dos valores.
STATUS_ATUAL	Codificação de rótulo para os valores numéricos

Fonte: Elaborada pela autora (2020)

A segunda etapa, **Treinamento**, consiste em executar os transformadores encontrados na etapa anterior à base de dados de treinamento. Esta base geralmente é composta por uma porcentagem dos dados originais oriundos do sistema educacional. Com os dados transformados, separam-se os dados de treinamento em duas partes. A primeira parte é passada a uma lista de preditores. A segunda parte, chamada de dados de validação, é passada aos preditores de forma que reportem seus desempenhos segundo alguma das métricas.

Para a avaliação dos desempenhos dos algoritmos, foram utilizadas as métricas da Matriz de Confusão que permite identificar a frequência das classes no modelo avaliado. E no formato gráfico, utilizou-se a análise de ROC (*Receiver Operating Characteristic*) e a Área sob a Curva ROC ou *Area Under the Curve* (ROC-AUC) para avaliar a qualidade do modelo de predição.

A terceira etapa, **Classificação**, consiste no fluxograma do sistema que será entregue. Este sistema conterà a inteligência através dos preditores treinados e a lista de tarefas de transformação. Portanto, ao apresentar dados não vistos pelo sistema, este os transformará e aplicará ao preditor mais apto ao problema avaliado e o resultado é disponibilizado para a avaliação e interpretação. Como forma de validar o método proposto da mineração dos dados educacionais e do aprendizado de máquina, foram realizados experimentos que estão detalhados no capítulo 05 e 06 por meio dos dados acadêmicos da Unitins.

### 3.3 Delimitação do estudo: Amostra da pesquisa

No contexto histórico a Universidade do Tocantins foi criada em 1990 e até o presente momento ocorreram várias mudanças de nome e no seu regime jurídico, ora autarquia, ora como fundação pública de direito privado. A mudança mais recente ocorreu por intermédio da Lei nº 3.124/2016, com a transformação do seu regime jurídico de fundação para autarquia de regime especial, que lhe foi garantida sua autonomia didático-científica, administrativa, financeira e patrimonial, passando a se chamar Universidade Estadual do Tocantins. Nesta

mesma lei houve a inclusão da reserva de 25% das vagas dos concursos de vestibulares, para estudantes provenientes da rede pública de ensino (UNITINS, 2019).

Atualmente a Unitins possui 5 câmpus universitários e num total de 15 cursos que estão distribuídos nos municípios de Araguatins, Augustinópolis, Dianópolis, Palmas e Paraíso do Tocantins (UNITINS, 2020a). Conforme a Lei de nº 3.124/2016 a previsão que seja implantado mais 3 (três) câmpus universitários, nos municípios de Guaraí, Taguatinga e Formoso do Araguaia (UNITINS, 2016).

O escopo da pesquisa abrangeu 12 cursos presenciais que estão distribuídos nos câmpus universitários de Araguatins, Augustinópolis, Dianópolis e Palmas. O período de análise dos dados foram de 9 anos, compreendidos entre os períodos letivos de 2010-2 até 2019-1. O câmpus de Paraíso do Tocantins não fez parte do escopo, pelo fato de ter iniciado as atividades no primeiro semestre de 2010. O quadro 9 apresenta os cursos presenciais analisados na pesquisa.

**Quadro 9** - Relação de cursos da Unitins por câmpus universitário

<b>Região</b>	<b>Câmpus</b>	<b>Curso</b>
Norte	Araguatins	Letras
		Pedagogia
	Augustinópolis	Ciências Contábeis
		Direito
		Enfermagem
	Sudeste	Dianópolis
Ciências Contábeis		
Direito		
Central	Palmas	Direito
		Engenharia Agrônômica
		Sistemas de Informação
		Serviço Social

Fonte: Elaborado pela autora com base em Unitins (2020a)

Em se tratando especificamente sobre os requisitos da perda de vínculo Institucional a Unitins regulamenta em seu regimento acadêmico, que está vigente desde 30/01/2020, na seção III “da Desvinculação” no art. 70. A perda de vínculo se compreenderá nas situações:

I – não efetuar matrícula no período estabelecido no calendário acadêmico da IES, caracterizando abandono de curso;

II – for reprovado por frequência e/ou média, em todas as disciplinas matriculadas no período letivo;

III – for reprovado três vezes em uma mesma disciplina;

IV – solicitar sua desvinculação do curso em qualquer momento;

V – não concluir o curso dentro do limite de tempo máximo fixado pelo Projeto Pedagógico do Curso.

§1o Perdido o vínculo com a Unitins, o acadêmico deixará de gozar dos direitos inerentes aos alunos regularmente matriculados.

§2o O coordenador do curso deverá ter conhecimento de todos os processos de desvinculação dos alunos do seu respectivo curso, logo após a sua instrução pela secretaria acadêmica, devendo ainda, manifestar-se sobre cada um deles.

§3o O coordenador do curso poderá após análise, manifestar-se pela não desvinculação de acadêmicos que estiverem matriculados no último semestre letivo do curso, nas hipóteses em que restarem somente as disciplinas de estágio e/ou TCC para integralização da estrutura curricular.

§4o A hipótese prevista no parágrafo anterior somente poderá ser aplicada uma única vez para cada acadêmico que a pleitear.

§5o A perda de vínculo prevista nesta seção não se confunde com a aplicação da sanção de desligamento, decorrente de processo administrativo disciplinar previsto na seção III, do capítulo X, deste regimento (UNITINS, 2020c).

A seguir, serão apresentados os trabalhos realizados como forma de subsidiar as discussões a respeito do tema e a realização dos experimentos baseados nos dados acadêmicos da Unitins, por meio das técnicas de inteligência artificial.

## CAPÍTULO DE LIVRO SUBMETIDO PARA EDITORA UNIVERSIDADE EdUFT

### EVASÃO ESCOLAR COM BASE EM PATENTES E REGISTROS DE SOFTWARES

**Resumo:** *A evasão escolar é um tema que vem sendo discutido com grande relevância, tanto no cenário nacional e internacional. O reflexo não é apenas para o aluno, mas para a sociedade como um todo, com consequências adversas, desde ordem econômica até ao aumento da desigualdade social do país. Diante dessa situação, o presente estudo objetivou avaliar os avanços tecnológicos na área da Evasão Escolar, uma vez que este problema não está restrito às Instituições de Ensino Brasileiras, fenômeno este que aflige também o cenário internacional. Para a realização da busca de patentes e registros de programas de computador, utilizou-se as bases de dados da Espacenet, Orbit.com e do Instituto Nacional de Propriedade Industrial (INPI). Por sua vez, os resultados mostram um pequeno número de tecnologias nacionais e internacionais, além disso, mesmo com esse número reduzido, muitas dessas tecnologias foram anuladas ou extintas. O cenário nos leva a acreditar, que este fenômeno ainda é pouco explorado, dificultando assim, a sua compreensão e causas.*

**Palavras-chaves:** *Abandono Escolar, Evasão Escolar; Evasão Escolar Estudantil.*

Autores: Leandra Cristina Cavina Piovesan Soares, Alexandre Rossini, Wosley Arruda, Francisco Gilson Rebouças Porto Junior, Rafael Lima de Carvalho



## 4 EVASÃO ESCOLAR COM BASE EM PATENTES E REGISTROS DE SOFTWARES

**Resumo:** *A evasão escolar é um tema que vem sendo discutido com grande relevância, tanto no cenário nacional e internacional. O reflexo não é apenas para o aluno, mas para a sociedade como um todo, com consequências adversas, desde ordem econômica até ao aumento da desigualdade social do país. Diante dessa situação, o presente estudo objetivou avaliar os avanços tecnológicos na área da Evasão Escolar, uma vez que este problema não está restrito às Instituições de Ensino Brasileiras, fenômeno este que aflige também o cenário internacional. Para a realização da busca de patentes e registros de programas de computador, utilizou-se as bases de dados da Espacenet, Orbit.com e do Instituto Nacional de Propriedade Industrial (INPI). Por sua vez, os resultados mostram um pequeno número de tecnologias nacionais e internacionais, além disso, mesmo com esse número reduzido, muitas dessas tecnologias foram anuladas ou extintas. O cenário nos leva a acreditar, que este fenômeno ainda é pouco explorado, dificultando assim, a sua compreensão e causas.*

**Palavras-chaves:** *Abandono Escolar, Evasão Escolar; Evasão Escolar Estudantil.*

### 4.1 Introdução

O combate à evasão escolar e a permanência por longo tempo dos alunos nas Instituições de Ensino Superior (IES) tornaram-se uns dos grandes desafios para a Educação Brasileira (INEP, 2019b). Tendo em vista o atual cenário, o estudo técnico aqui apresentado tem por objetivo avaliar o cenário mundial, no que diz respeito às soluções registradas que lidam com a Evasão Escolar. Este levantamento será feito por meio de buscas de informações tecnológicas e inovadoras, cadastradas em bases de patentes e de registros de programas de computador. De acordo com (CARVALHO e colab., 2014), bases de patentes são grandes armazenadores de dados e importantes fontes para as organizações, na busca de inovações que são geradas a partir da Pesquisa e Desenvolvimento (P&D).

Para os autores (PARANHOS e colab., 2018) a proteção de uma carta de patente é gerada pelo fruto de um processo de pesquisa e desenvolvimento, muitas vezes demorado e oneroso. Pesquisas e desenvolvimentos de novos produtos envolvem esforços, custos e na maioria das vezes, são investimentos altos. A proteção deste produto significa uma forma de se resguardar contra os competidores, seja por cópias ou vendas não autorizadas, além disso, a proteção é considerada um valioso e fundamental mecanismo jurídico, para que este esforço se converta em investimentos seguros e rentáveis (SABINO, 2007).

O acesso às informações tecnológicas também por ser vistas por meio dos registros de Programas de Computador, os quais protegem a parte intelectual por meio do código fonte,

diferentemente da patente que protege o produto. Segundo INPI (2019a) com o crescimento do valor atribuído aos *softwares*, houve um aumento nas buscas de como proteger seus ativos intangíveis, por parte das organizações. Mesmo não sendo obrigatório, o registro de programa de computador passou a ser um dos mecanismos utilizados pelas indústrias, no que diz respeito a cópias não autorizadas, garantindo maior segurança jurídica. De acordo com (AMPARO e colab., 2012) a prospecção tecnológica, de maneira geral, visa a minimização de incertezas para a tomada das decisões estratégicas. Uma vez que a prospecção, pode contribuir de forma mais eficaz e econômica para a busca de soluções tecnológicas.

Neste contexto, cada vez mais as organizações têm buscado fontes de informações tecnológicas, para melhor conhecer o seu ambiente de atuação. Este conhecimento permite uma decisão mais fundamentada na definição das estratégias para o desenvolvimento de novos produtos e processos. No entanto, a capacidade de inovação é requisito primordial para o ambiente globalizado, onde os avanços tecnológicos abriram um vasto campo para pesquisas, permitindo assim o surgimento de novas relações entre organizações e os países onde atuam.

No presente estudo, serão contextualizados o cenário nacional sobre a Evasão Escolar, bem como as definições dos registros de programas de computadores e patentes. Em seguida, serão apresentados os procedimentos metodológicos adotados nesta pesquisa. Por final, serão apresentados os resultados das buscas, juntamente com as discussões do estudo realizado.

#### **4.2 O cenário Nacional da Evasão Escolar**

A evasão escolar de forma comum é um dos problemas que preocupam as Instituições de Ensino (FILHO e colab., 2007). Em um panorama nacional, estudos anteriores realizados na Universidade Federal do Paraná (UFPR) e na Pontifícia Universidade Católica do Paraná, já indicavam a existência da evasão nos cursos universitários. Segundo (PAREDES, 1994) este fenômeno é muito maior do que a percepção geral que as universidades têm sobre ele. As instituições encaram como se fosse algo comum e normal e que fazem parte do cenário universitário mundial. PAREDES (1994), considera que a sub-avaliação e a falta de interesses e investigações sobre o este problema acabam gerando decisões administrativas impróprias a produtividade geral dos cursos. Sob o ponto de vista da Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras

(CEUPB), a Evasão Escolar é um fenômeno complexo e comum as IES do mundo contemporâneo e que vem sendo influenciado por diversas variáveis, nos quais provocam a necessidade de estudos e análises sobre tal tema (BRASIL, 1996b).

No entanto, estudos realizados pela Associação de Mantenedoras de Ensino Superior (ABMES), defende sobre a necessidade uma política governamental orientada a qualidade do ensino, com uso dos recursos públicos e privados, sendo conduzidos para a promoção de processos e análises direcionadas para realizações de ações. Onde os motivadores deveriam ser os órgãos de governo gestores ou os fiscalizadores das IES públicas e privadas (HORTA, 2012). E para um melhor entendimento sobre este fenômeno, os autores (KIRA, 1998; GAIOSO, 2005; (BAGGI e colab., 2011), definem que a Evasão Escolar é a interrupção do ciclo de estudos, em qualquer nível de ensino. Em outro estudo realizado por (SCHARGEL e SMINK, 2002) a investigação para este fenômeno deve-se ser dividido em categorias de causas da evasão, sendo: psicológicas, sociológicas, organizacionais, interacionais e as econômicas. Por fim, (FILHO e colab., 2007) ressalta que a problemática da evasão escolar é de ordem internacional e afeta diretamente o desempenho dos sistemas educacionais.

Diante a várias ponderações sobre esta problemática do ensino superior brasileiro, que é considerado um dos principais desafios para o sistema educacional brasileiro. O cenário tecnológico computacional atual possui ferramentas extraordinárias dentro das áreas de Inteligência Artificial, Aprendizado de Máquina e Mineração de Dados. Neste sentido, uma vez que tais ferramentas se tornam cada vez mais acessíveis, o uso de tecnologias inovadoras são cada vez mais necessárias para as Instituições de ensino. Principalmente, em tarefas de análise e extração de conhecimentos a partir de seus dados, priorizando a geração de soluções eficazes para a tomada de decisão. Outrossim, justifica-se o conhecer das tecnologias existentes para lidar com problemas da evasão.

### **4.3 Sobre os Registros de Programas de Computador e Patentes**

A formalização do registro de programas de computador, garante ao autor o direito de reivindicar contra quaisquer alterações e cópias não autorizadas, tendo como particularidade a sua abrangência que é internacional, garantida através da Convenção de Berna (BRASIL, 1994). De acordo com a *World Intellectual Property Organization* (WIPO), a Convenção de Berna é um dos tratados mais antigos, criado em 1886 para atender uma necessidade de uniformização do sistema internacional, com foco em proteção de obras literárias e artísticas.

Entre as décadas de 70 e 80 iniciaram-se as discussões sobre os tipos de proteções de *softwares*, sob a Lei de direitos autorais, patentes ou a proteção *sui generis*. Mas foi em 1985 que a WIPO e a Organização das Nações Unidas para Educação, Ciência e Cultura (UNESCO), formaram um comitê e decidiram tornar os direitos autorais como a proteção utilizada para os programas de computador (WIPO, 2008).

De acordo com (BRASIL, 1998a; 1998b), o programa de computador é regido pela Lei nº 9.610/98 com proteção sobre obras intelectuais. Entretanto, os programas de computador são objetos de legislação específica, a chamada “Lei de *Software*” de nº 9.609/98, que concede a proteção da propriedade intelectual aos programas de computador, como obras literárias e pela legislação de direitos autorais e conexos. Um programa de computador pode ser definido:

Programa de computador é a expressão de um conjunto organizado de instruções em linguagem natural ou codificada, contida em suporte físico de qualquer natureza, de emprego necessário em máquinas automáticas de tratamento da informação, dispositivos, instrumentos ou equipamentos periféricos, baseados em técnica digital ou análoga, para fazê-los funcionar de modo e para fins determinados (BRASIL, 1998a).

De acordo com (TIGRE e MARQUES, 2009); INPI, 2019a) a proteção do direito autoral está relacionada à proteção do código fonte, por meio da "forma de expressão da ideia" e não à "aplicação da ideia" que o *software* executa. Para cada nova versão desenvolvida requer um novo registro e não há limitações de quantitativos de registros depositados sobre um mesmo *software*. O registro têm a validade por 50 anos, a partir do dia 1º de janeiro do ano subsequente da sua publicação, onde a garantia de propriedade é no ato da sua criação, independente do registro. Além disso, não requerer exame para o registro e a proteção é compreendida em 175 países, que são signatários da Convenção de Berna (INPI, 2018).

As proteções conferidas para as Patentes de Invenção ou Modelo de Utilidade (MU) é de propriedade industrial, regulamentada pela Lei de nº 9.279/96 dos direitos e obrigações relativos à propriedade industrial (BRASIL, 1996a). Segundo (BARBOSA, 2003; INPI, 2019b;) a patente de invenção ou modelo de utilidade é um título provisório outorgado pelo Estado aos inventores para a exploração de uma tecnologia. A patente permite ao inventor (ou detentor do registro) o direito de impedir que terceiros, sem seu consentimento, façam uso, coloque a venda ou importe o produto patentado. Em compensação, o inventor revela por

meio de documentos os conhecimentos do seu invento. Além disso, tem que se levar em consideração não a ideia como foi demonstrada, mas a sua aplicação prática. De acordo com (ANDRADE e colab., 2009) a invenção por patente deverá atender os requisitos de atividade inventiva, inovadora e com aplicação industrial.

Para os autores (RUSSO e colab., 2012) as patentes possuem validade por um período limitado, normalmente de 20 anos e o MU de 15 anos, contados a partir da data do depósito. A expedição é feita de acordo com a legislação de cada país, no caso do Brasil é efetuado por um órgão governamental (INPI) e nos demais países por escritórios de patentes. A proteção tem a abrangência territorial, onde ela foi concedida e para que seja solicitado a proteção em outros países, é necessário que o documento seja traduzido para o idioma do país/região. Além disso, é necessário a nomeação de um procurador para representação naquele determinado país. O depósito em diferentes países pode ser simplificado, por meio do Tratado de Cooperação de Patentes (PCT). Este tratado conta com 152 países signatários, entre eles o Brasil e esta proteção se torna mais econômica, quando for pedida em vários países (INPI, 2019b).

Assim sendo, no Brasil prevalecem dois tipos de proteções, do direito autoral e da propriedade industrial. Da propriedade industrial são às patentes de invenções para produtos/processos e MU. E no campo dos direitos autorais os *softwares* estão classificados nesta categoria. Autores como (SEGUNDO e TELES, 2019) argumentam que os *softwares* possuem evidências para a aplicação industrial, onde são empregados esforços inventivos no formato de novas técnicas e produtos. Onde são agregados valores econômicos e voltados para soluções de problemas. Este tipo de criação habitual se dá na forma de patentes, já que salvaguarda os requisitos de novidade, aplicação industrial e atividade inventiva. Já para os autores (TIGRE e MARQUES, 2009) o patenteamento de *softwares* implicaria em pagamentos conjuntos de várias licenças, já que os sistemas são construídos por vários subsistemas. Desta forma, a comercialização deste produto resultaria em custos mais elevados para a sociedade. Além disso, as patentes de *softwares* geraria um gargalo no aumento dos pedidos, onde as pequenas empresas não teriam condições de competir com as grandes empresas, já que estas grandes corporações possuem advogados especializados com condições de gerar um número expressivo de pedidos.

#### 4.4 Metodologia

De acordo com (LEME e DE MELLO, 2019), os estudos de vigilância tecnológica (VT) permitem a análise do desenvolvimento do estado da arte da tecnologia pesquisada, elucidando oportunidades e possibilitam prospecção na área de interesse. Neste sentido, para efetuar a pesquisa das Patentes e dos Registros de Programas de Computador, foram utilizadas as bases de dados: Orbit.com, Espacenet e INPI. Foram escolhidas 2 bases internacionais e 1 nacional. A base de dados Orbit.com é um sistema comercial para busca, seleção e análise de patentes (AXONAL, 2016). A Espacenet é uma base gratuita, mantida pelo escritório Europeu de Patentes (EPO) e possui indexação de mais de 100 milhões de documentos de patentes (ESPACENET, 2020).

O INPI é uma autarquia Brasileira, fundada em 1970, vinculada ao Ministério da Economia e tem como missão o estímulo à inovação e competitividade, com fomento ao desenvolvimento tecnológico por meio da proteção da propriedade industrial. A base de buscas é gratuita e de acordo com relatório de Estatísticas Preliminares do ano de 2019, o acumulado de patentes entre 2013 a 2019 foi de 28.318 pedidos de patentes. Já em relação aos programas de computadores, neste mesmo período, somaram um montante de 3.048 depósitos residentes no Brasil (INPI, 2020c).

As buscas e recuperação dos documentos considerou uma janela de tempo no intervalo entre 2004 a 2019. As palavras-chaves utilizadas na busca das patentes e dos registros de programas de computador foram baseadas por meio de consultas em artigos científicos. A tabela 1, apresenta as expressões utilizadas com seus respectivos quantitativos de depósitos de patentes e registros de programas de computador.

**Tabela 1** - Expressões utilizadas nas buscas de Patentes e Programas de Computador

Busca	Expressão	Qtde de Patentes	Qtde de Registro de Programas de Computador
01	Título, Resumo e Reivindicação= <b>School Evasion</b>	1	0
02	Título, Resumo e Reivindicação= <b>School Dropout</b>	66	0

03	Contenha: a palavra aproximada: <b>'Evasão'</b> no Título Programa	0	4
04	Contenha: <b>"Evasão escolar"</b> no Título do Programa	0	1
05	Contenha: <b>"Evasão escolar"</b> no Título	1	0
06	Contenha: <b>"Evasão escolar"</b> no Resumo	3	0
07	Título, Resumo, Objeto da Invenção, Reivindicações Independentes e Texto Completo: <b>Evasão Escolar</b>	7	0
08	Título, Resumo, Objeto da Invenção, Reivindicações Independentes e Texto Completo: <b>School Dropout</b>	25	0
09	Título, Resumo, Objeto da Invenção, Reivindicações Independentes e Texto Completo: <b>School+evasion+student</b>	7	0
<b>Total</b>		<b>110</b>	<b>5</b>

Fonte: Elaborado pelos autores com base em (INPI, 2020a, 2020b; Questel, 2020; Espacenet, 2020)

Como as buscas foram efetuadas em 3 (três) bases diferentes, foi necessário o uso de palavras-chaves em inglês. O total de documentos recuperados foram de 110 (cento de dez) patentes e 5 (cinco) registros de programas de computador. O próximo passo, foi a remoção dos registros duplicados e daqueles que não faziam parte do tema de Evasão Escolar, este número foi reduzido para 09 (nove) patentes e 4 (quatro) registros de programas de computador. A partir destes números, foi criada uma base de dados para a análise dos resultados.

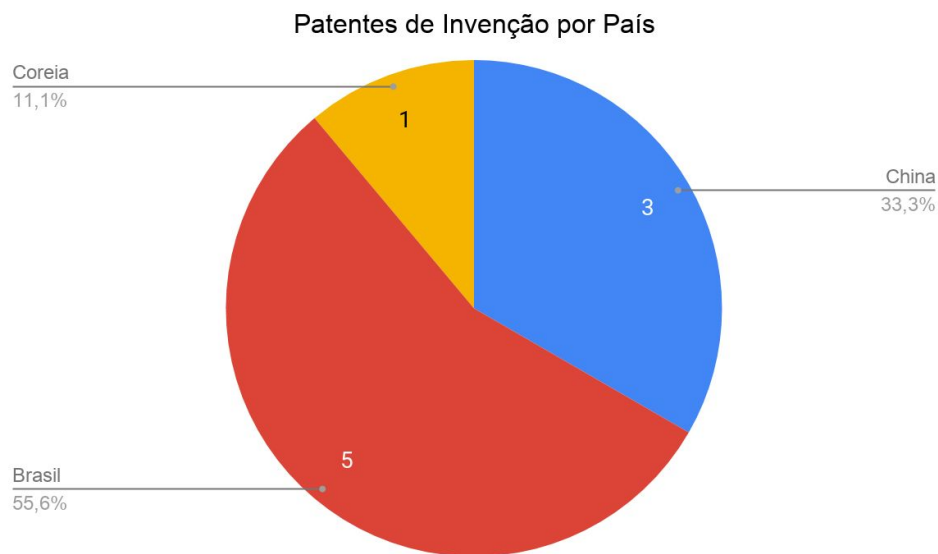
#### 4.5 Resultados e Discussão

Por meio da análise dos documentos de patentes e registros de programas de computador, foi possível realizar um mapeamento para melhor conhecer o cenário nacional e mundial, bem como os países que detêm as tecnologias voltadas para a Evasão Escolar e a posição do Brasil comparado aos demais países.

O gráfico 1 apresenta o ranking do número de patentes para tecnologias de Evasão Escolar por país, independente se as patentes estão ativas ou não. Observa-se que o Brasil é o

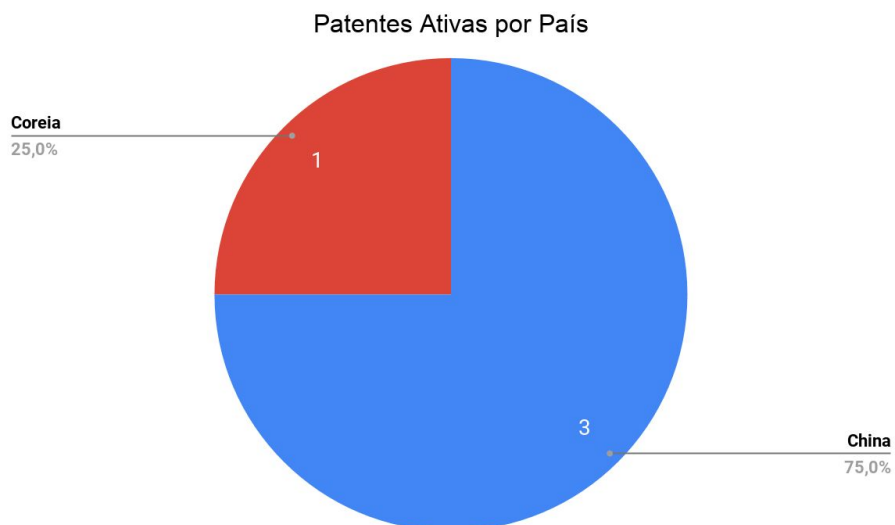
maior depositário de patentes com o total de 5 (cinco) tecnologias, em seguida vem a China com 3 (três) e por último a Coreia com 1 (um). Ao aplicar o filtro para exibir apenas patentes ativas, esse número é reduzido para 4 (quatro). O Brasil deixa de aparecer na relação, ficando apenas a China com 3 (três) patentes e a Coreia com 1 (um). O gráfico 2, apresenta os países com patentes ativas.

**Gráfico 1 - Quantitativo de Patentes por País**



Fonte: Elaborado pelos autores com base em (INPI, 2020a; Questel, 2020; Espacenet, 2020)

**Gráfico 2 - Países com patentes ativas**

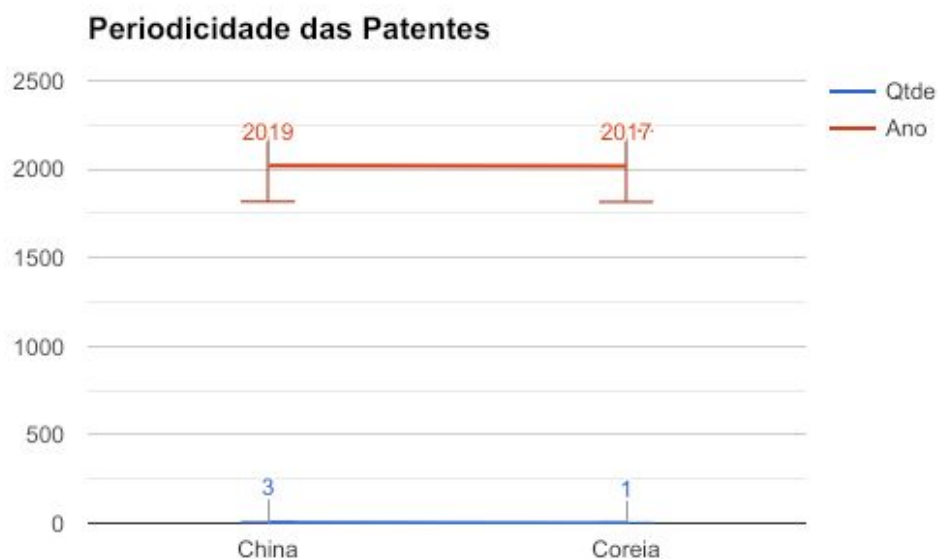




Fonte: Elaborado pelos autores com base em (INPI, 2020a; Questel, 2020; Espacenet, 2020)

Em relação aos quantitativos de depósitos de patentes por ano, observa-se que o ano de 2019 teve o maior número de depósitos. A China fez o depósito de 3 (três) patentes e a Coreia em 2017, apenas 1 (um). O gráfico 3 apresenta este quantitativo.

**Gráfico 3 - Depósito de Patentes por Ano**



Fonte: Fonte: Elaborado pelos autores com base em (INPI, 2020a; Questel, 2020; Espacenet, 2020)

Quanto ao número de registros de programas de computador, são 5 (cinco) no total o primeiro foi no ano de 2014 e o último em 2019 (INPI, 2020b). Nota-se que este número é bem reduzido e somente o Brasil é detentor dos direitos para tecnologias voltados para este fenômeno. Mesmo sendo um processo de registro menos burocrático, ágil e com custos mais acessíveis, não foram encontrados registros de outros países na base do INPI.

#### **4.6 Considerações Finais**

Neste capítulo foi possível entender os esforços empreendidos para mitigar a Evasão Escolar no Ensino Superior, por meio de estudos que buscam definições e causas que levam a solução do problema. Pode-se dizer que busca-se por meio dos avanços tecnológicos, soluções que auxiliem as IES na tomada de decisão. Foi apresentado também, o histórico

legislação, definição e formato da proteção dos Programas de Computador. Este tipo de direito é caracterizado como obras intelectuais, por meio da Lei dos direitos autorais. Já, para as Patentes de Invenção e Modelos de Utilidades a proteção é relacionada a Propriedade Industrial. Diferentemente dos programas de computador, este tipo de proteção requer um produto ou processo associado com aplicação comercial. Além disso, apresentou-se as discussões, motivos e implicações sobre o patenteamento dos *softwares*. A maioria dos países, inclusive o Brasil adotaram a proteção intelectual para os programas de computador.

Para os números recuperados de patentes e registros de programas de computador, extraídos das bases de dados, foi apresentado um panorama simplificado sobre as tecnologias para a Evasão Escolar em um período de 15 anos. Nota-se que os quantitativos recuperados tanto de programas de computador e patentes, são poucos significativos para o quão grande e complexo é este fenômeno.

Assim, é possível perceber que os avanços tecnológicos ainda são insuficientes, requer mais estudos e desenvolvimento de soluções de softwares e/ou patentes para este campo. Algumas hipóteses podem ser consideradas para este número reduzido, como a recuperação dos documentos não atender a nomenclatura de registro das proteções, as tecnologias desenvolvidas pelas próprias instituições, não são registradas nos órgãos responsáveis, seja ela por falta de conhecimento do processo ou até mesmo por questões financeiras. E por fim, os programas de computador não requer registro, o direito do autor é garantido independentemente da proteção junto ao INPI.

## Referências

AMPARO, Keize Katiane dos Santos e colab. Estudo de caso utilizando mapeamento de prospecção tecnológica como principal ferramenta de busca científica. *Perspectivas em Ciência da Informação*. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s1413-99362012000400012>>. , 2012

ANDRADE, Elvira e colab. Propriedade Intelectual em Software: o que podemos apreender da experiência internacional? *Revista Brasileira de Inovação*. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.20396/rbi.v6i1.8648940>>. , 2009

AXONAL. Axonal Consultoria Tecnológica (2016). Disponível em: [https://www.axonal.com.br/arquivos/PDF/Orbit\\_Visao\\_Geral\\_Sistema\\_PARTES\\_1\\_a\\_3\\_BU\\_SCA\\_VISUALIZACAO\\_SELECAO.pdf](https://www.axonal.com.br/arquivos/PDF/Orbit_Visao_Geral_Sistema_PARTES_1_a_3_BU_SCA_VISUALIZACAO_SELECAO.pdf). Acesso em: 13/04/2020

BAGGI, Cristiane Aparecida Dos Santos e DOS SANTOS BAGGI, Cristiane Aparecida e LOPES, Doraci Alves. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. Avaliação: Revista da Avaliação da Educação Superior (Campinas). [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s1414-40772011000200007>>. , 2011

BARBOSA, Denis Borges (2003). Uma introdução a Propriedade Intelectual. Disponível em: <http://www.denisbarbosa.addr.com/arquivos/livros/umaintro2.pdf>. Acesso em: 22/04/2020

BRASIL. Decreto nº 1.355, de 30 de dezembro de 1994. Promulgo a Ata Final que Incorpora os Resultados da Rodada Uruguai de Negociações Comerciais Multilaterais do GATT. DF, dez 1994.

BRASIL. Lei nº 9.279, de maio de (1996a). Regula direitos e obrigações relativos à propriedade industrial. DF, mai 1996.

BRASIL. Ministério da Educação. Secretaria de Educação Superior (1996b). Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. Acesso em: 11/04/2020. Disponível em: [http://www.andifes.org.br/wp-content/files\\_flutter/Diplomacao\\_Retencao\\_Evasao\\_Graduacao\\_em\\_IES\\_Publicas-1996.pdf](http://www.andifes.org.br/wp-content/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf). Acesso em: 15/03/2020

BRASIL, Lei nº 9.609, de 19 de fevereiro de (1998a). Dispõe sobre a proteção da propriedade intelectual de programa de computador, sua comercialização no País, e dá outras providências. DF, fev 1998

BRASIL, Lei nº 9.610, de 19 de fevereiro de (1998b). Altera, atualiza e consolida a legislação sobre direitos autorais e dá providências. DF, fev 1998

CARVALHO, Alexsandro e STOROPOLI, João e QUONIAM, Luc. Prospecção de Patentes para a Solução Sustentável de Problema da Indústria da Construção: O Espaçador de Concreto. Revista Inovação, Projetos e Tecnologias. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.5585/iptec.v2i1.21>>. , 2014

ESPACENET. Espacenet Patent search (2020). Disponível em: <https://worldwide.espacenet.com/>. Acesso em: 02/04/2020

FILHO, Roberto Leal Lobo e. Silva e colab. A evasão no ensino superior brasileiro. Cadernos de Pesquisa. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s0100-15742007000300007>>. , 2007

GAIOSO, N. P. L. O fenômeno da evasão escolar na educação superior no Brasil. 2005. 75 f. Dissertação (Mestrado em Educação) Programa de Pós-Graduação em Educação da Universidade Católica de Brasília, Brasília, DF, 2005.

HORTA, Cecília Eugenia Rocha. Associação Brasileira de Mantenedoras de Ensino Superior (2012). Evasão no ensino superior brasileiro. Disponível em: <https://abmes.org.br/arquivos/publicacoes/Cadernos25.pdf>. Acesso em: 27/04/2020

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019). Sem desistências, número de graduados poderia dobrar no Brasil. Disponível em: [http://portal.inep.gov.br/artigo/-/asset\\_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206](http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206). Acesso em 08/10/2019.

INPI. Instituto Nacional da Propriedade Industrial (2018). Perguntas Frequentes – Programas de Computador. Disponível em: <http://www.inpi.gov.br/servicos/perguntas-frequentes-paginas-internas/perguntas-frequentes-programa-de-computador#faq1.0>. Acesso em: 07/04/2020

INPI. Instituto Nacional da Propriedade Industrial (2020a). Consulta à Base de Dados do INPI. Disponível em: <https://gru.inpi.gov.br/pePI/jsp/patentes/PatenteSearchBasico.jsp>. Acesso em: 06/03/2020

INPI. Instituto Nacional da Propriedade Industrial (2020b). Consulta à Base de Dados do INPI - Base Programas. Disponível em: <https://gru.inpi.gov.br/pePI/jsp/programas/ProgramaSearchBasico.jsp>. Acesso em: 06/03/2020

INPI. Instituto Nacional da Propriedade Industrial (2020c). Boletim Mensal de Propriedade Industrial - Estatísticas Preliminares. Disponível em: [http://www.inpi.gov.br/sobre/estatisticas/arquivos/publicacoes/boletim\\_jan\\_2020.pdf](http://www.inpi.gov.br/sobre/estatisticas/arquivos/publicacoes/boletim_jan_2020.pdf). Acesso em 13/04/2020

INPI. Instituto Nacional de Propriedade Industrial (2019a). Manual do Usuário para o Registro Eletrônico de Programas de Computador. Disponível em: <http://www.inpi.gov.br/menu-servicos/programa-de-computador/programa-de-computador-manual-completo>. 44 f.; il.; anexos. Versão 1.8.5. Data: 07/05/2019. Acesso em: 23/04/20

INPI. Instituto Nacional da Propriedade Industrial (2019b). Perguntas Frequentes - Patentes. Disponível em: <http://www.inpi.gov.br/servicos/perguntas-frequentes-paginas-internas/perguntas-frequentes-patente#patente>. Acesso em: 07/04/ 2020

INPI. Instituto Nacional da Propriedade Industrial (2019c). O INPI. Disponível em: <http://www.inpi.gov.br/sobre/estrutura>. Acesso em: 13/04/2020

KIRA, Luci Frare. A evasão no ensino superior: o caso do curso de pedagogia da Universidade Estadual de Maringá (1992-1996). 1998. 106 f. Dissertação (Mestrado em Educação) – Programa de Pós-graduação em Educação da Universidade Metodista de Piracicaba, Piracicaba, 1998.

LEME, Joao Vitor e DE MELLO, Flavia Luciane Consoni. Tecnologias para os veículos a célula a combustível. Revista dos Trabalhos de Iniciação Científica da UNICAMP. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.20396/revpibic262018685>>. , 2019

PARANHOS, Rita D. E. Cassia Santos e DE CASSIA SANTOS PARANHOS, Rita e RIBEIRO, Núbia Moura. IMPORTÂNCIA DA PROSPECÇÃO TECNOLÓGICA EM BASE EM PATENTES E SEUS OBJETIVOS DA BUSCA. Cadernos de Prospecção. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.9771/cp.v12i5.28190>>. , 2018

PAREDES, Alberto Sánchez (1994). A Evasão do Terceiro Grau em Curitiba. Disponível em: <http://nupps.usp.br/downloads/docs/dt9406.pdf>. Acesso em: 27/04/2020

QUESTEL. Data Coverage (2020). Disponível em: <https://www.questel.com/data-coverage/>. Acesso em 14/03/2020

RUSSO, Suzana Leitão, SILVA, Gabriel Francisco da, NUNES, Maria Augusta Silveira Netto (2012). Capacitação em inovação tecnológica para empresários. Disponível em: <http://www.api.org.br/bancodearquivos/uploads/57410-livro-capacite-v2.pdf#page=56>. Acesso em: 24/04/2020

SABINO, Luciana Shicasho (2007). Caracterização da Proteção às patentes como estímulo do desenvolvimento econômico. Disponível em: <https://bdtd.ucb.br:8443/jspui/bitstream/123456789/387/1/Texto%20completo.pdf>. Acesso em: 24/04/2020

SCHARGEL, F. P.; SMINK, J. Estratégias para auxiliar o problema de evasão escolar. Tradução de Luiz Frazão Filho. Rio de Janeiro: Dunya, 2002

SEGUNDO, Gesil Sampaio Amarante, TELES, Eduardo Oliveira (2019). PROFNIT. Conceitos e aplicações de propriedade intelectual; V. 2. Disponível em: [http://www.profnit.org.br/wp-content/uploads/2019/11/PROFNIT-Serie-Conceitos-e-Aplicacao-A7-B5es-de-Propriedade-Intelectual-Volume-II-PDF\\_compressed.pdf](http://www.profnit.org.br/wp-content/uploads/2019/11/PROFNIT-Serie-Conceitos-e-Aplicacao-A7-B5es-de-Propriedade-Intelectual-Volume-II-PDF_compressed.pdf). Acesso em: 28/04/2020

TIGRE, Paulo Bastos e MARQUES, Felipe Silveira. Apropriação tecnológica na economia do conhecimento: inovação e propriedade intelectual de software na América Latina. Economia e Sociedade. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s0104-06182009000300005>>. , 2009

WIPO. Intellectual property handbook, 2ª edição 2004 e reimpresso 2008. Disponível em: [https://www.wipo.int/edocs/pubdocs/en/intproperty/489/wipo\\_pub\\_489.pdf](https://www.wipo.int/edocs/pubdocs/en/intproperty/489/wipo_pub_489.pdf). Acesso em: 06 de abr. 2020

## **ARTIGO PUBLICADO NA REVISTA HUMANIDADES & INOVAÇÃO**

Revista Humanidades & Inovação ISSN: 2358-832

<http://diadorim.ibict.br/handle/1/1187>

### **APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA EM UM CONTEXTO ACADÊMICO COM FOCO NA IDENTIFICAÇÃO DOS ALUNOS EVADIDOS E NÃO EVADIDOS**

**Resumo:** A evasão escolar é um dos principais problemas causadores de prejuízos às Instituições de Ensino Superior e a utilização de modelos de predição podem subsidiar decisões para minimização dos prejuízos. Nesse contexto, este trabalho avalia se é possível empregar algoritmos de aprendizado de máquina para gerar modelar o padrão de evasão, a partir de dados de registro acadêmico. Esta hipótese foi validada através de um estudo de caso, usando os dados acadêmicos da Universidade Estadual do Tocantins. Os resultados alcançados pelos experimentos indicaram que a metodologia adotada neste trabalho foi capaz de classificar com elevado grau de confiança os alunos em situação de evasão e de não evasão.

**Palavras-chaves:** Evasão Escolar. Mineração de Dados Educacionais. Aprendizagem de Máquina. Inovação em Gestão Educacional.

Autores:

Leandra Cristina Cavina Piovesan Soares

Robson Aparecido Ronzani

Rafael Lima de Carvalho

Alexandre Tadeu Rossini da Silva

## 5 APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA EM UM CONTEXTO ACADÊMICO COM FOCO NA IDENTIFICAÇÃO DOS ALUNOS EVADIDOS E NÃO EVADIDOS

APPLICATION OF MACHINE LEARNING TECHNIQUES IN AN ACADEMIC CONTEXT WITH A FOCUS ON IDENTIFYING DROPOUT AND DROPOUT STUDENTS

Aceito em 8 de maio de 2020

**Resumo:** A evasão escolar é um dos principais problemas causadores de prejuízos às Instituições de Ensino Superior e a utilização de modelos de predição podem subsidiar decisões para minimização dos prejuízos. Nesse contexto, este trabalho avalia se é possível empregar algoritmos de aprendizado de máquina para gerar modelar o padrão de evasão, a partir de dados de registro acadêmico. Esta hipótese foi validada através de um estudo de caso, usando os dados acadêmicos da Universidade Estadual do Tocantins. Os resultados alcançados pelos experimentos indicaram que a metodologia adotada neste trabalho foi capaz de classificar com elevado grau de confiança os alunos em situação de evasão e de não evasão.

**Palavras-chave:** Evasão Escolar. Mineração de Dados Educacionais. Aprendizagem de Máquina. Inovação em Gestão Educacional.

**Abstract:** School dropout is one of the main problems that cause losses to Higher Education Institutions and the use of prediction models can subsidize decisions to minimize losses. In this context, this work evaluates whether it is possible to employ machine learning algorithms to generate modeling the evasion pattern, based on academic record data. This hypothesis was validated through a case study, using academic data from the State University of Tocantins. The results achieved by the experiments indicated that the methodology adopted in this work was able to classify students in situations of evasion and non-evasion with a high degree of confidence.

**Keywords:** School Dropout. Educational Data Mining. Learning Machine. Innovation in Educational Management

### Introdução

O problema de evasão, ou abandono escolar, tem sido muito discutido no meio acadêmico (GUIMARAES et al, 2019; GOMES et al., 2019; SABBATINI, 2015; SANTOS JUNIOR; REAL, 2019; ADACHI, 2017; CARVALHO, 2017; SOUZA, 2017). Esse fenômeno afeta todos os níveis educacionais, desde os anos iniciais até ao ensino superior, independentemente do tipo de instituição, seja ela pública ou

privada, e a sua consequência impacta no desenvolvimento humano (CUNHA e MOSORINI, 2013). De acordo com (Kira, 1998; Gaioso, 2005; BAGGI e colab., 2011), a evasão escolar é definida pela interrupção do ciclo de estudos, em qualquer nível de ensino. Para Kira (1998), a evasão escolar é considerada como fuga ou perda dos alunos antes da conclusão do curso. Diversos fatores podem levar ao abandono escolar e nem sempre o motivo está relacionado ao mau desempenho acadêmico. Assim, na busca de uma melhor compreensão para esse fenômeno, surgem conceitos e estudos sobre a evasão escolar. De acordo com Souza (2008), a motivação da evasão não está apenas relacionada ao âmbito acadêmico, a causalidade pode advir de aspectos pessoais, sociais e ambientais. Segundo (BAGGI e colab., 2011), o abandono escolar requer uma análise histórica, pois a realidade dos níveis anteriores de ensino podem influenciar no abandono de um curso superior. Nesse contexto, percebe-se o quanto esse fenômeno é complexo uma vez que há vários aspectos envolvidos e requer estudo aprofundado. Estudar a evasão escolar no ensino superior a fim de extrair informações/conhecimentos que possam ser utilizados para minimizá-lo é o objetivo deste trabalho. Assim, assume-se como hipótese a ser verificada que técnicas de Aprendizado de Máquina podem ser aplicadas de maneira satisfatória em mineração de dados educacionais. Para verificar a hipótese, foram utilizados, como estudo de caso, os dados acadêmicos da Universidade Estadual do Tocantins (Unitins). Por meio de técnicas computacionais de Aprendizado de Máquina, os dados foram investigados e analisados a fim de prever a evasão antes de um aluno abandonar o curso. Aprendizado de Máquina (AM) é um ramo da Inteligência Artificial que compreende o desenvolvimento de técnicas e sistemas capazes de adquirirem conhecimento de maneira automatizada. A construção de um sistema de AM é baseada em experiências acumuladas através da solução de problemas anteriores (MONARD e BARANAUSKAS, 2003). Deste modo, a partir de dados acadêmicos da Unitins, esta pesquisa utilizou AM para produzir um classificador preditivo de alunos evadidos e não evadidos. Para melhor compreensão deste trabalho, será apresentado (a): o contexto da evasão escolar no ensino superior, em especial da Unitins; a relação de procedimentos metodológicos utilizados nesta pesquisa; o conjunto de testes e seus resultados, seguidos de uma discussão sobre eles.



## O contexto da evasão escolar no ensino superior

O fenômeno da evasão escolar e o estímulo à permanência por longo tempo dos alunos nas Instituições de Ensino Superior (IES) é um dos grandes desafios para a educação brasileira (INEP, 2019b). Para a Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras (CEUB), este estudo é complexo e comum as IES do mundo contemporâneo e que vem sendo influenciado por diversas variáveis, nos quais provocam a necessidade de estudos e análises sobre tal tema (BRASIL, 1996b). Para (FILHO e colab., 2007) a evasão escolar é um problema de ordem internacional que afeta os sistemas educacionais. No que tange ao setor público, os recursos investidos tem seu retorno comprometido por causa da evasão escolar, uma vez que gera ociosidade de vagas, equipamentos e espaços físicos; para o setor privado há perda de receitas. Além disso, a evasão escolar causa desperdícios sociais, acadêmicos e econômicos. Segundo a Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras (CEUB), a definição desse fenômeno é como a saída definitiva do aluno de seu curso de origem, sem concluí-lo (BRASIL, 1996b). Para melhor entendimento sobre a distinção de formas de evasão escolar, a CEUB apresenta, por meio do quadro 1 (um) os tipos de evasão.

**Quadro 1** - Distinção sobre conceitos de Evasão Escolar

Tipo	Descrição
Evasão do Curso	É quando o estudante é desligado do curso superior em situações diversas tais como: abandono (deixa de matricular-se), desistência (oficial), transferência ou reopção (mudança de curso), exclusão por norma institucional.
Evasão da Matrícula	É quando o estudante desliga-se da instituição na qual está matriculado.
Evasão do Sistema	Quando o estudante abandona de forma definitiva ou temporária o ensino superior.

Fonte: Elaborado pelos autores com base em BRASIL (1996b)

Os conceitos apresentados no quadro 1 (um) são importantes para compreender os trabalhos que mapearam as causas que levam os alunos a abandonarem seus cursos. Um estudo realizado por (COLPANI, 2019) identificou que o Pará foi o Estado com maior índice de evasão escolar no Brasil em 2017. O

autor utilizou os dados públicos do Censo no seu estudo e os resultados apontaram que a média de evasão nas escolas é de 20%. Deste percentual, entre 38% a 65% são dos alunos estão com idade acima recomendada pela Lei 9.394/96, que é de 18 anos (BRASIL, 1996a). Assim, o indicador da taxa de distorção de série/idade foi a variável associada com a evasão escolar. Souza (2008) realizou um estudo nos cursos de Engenharias da Universidade Federal Fluminense (UFF) e identificou que 32% das evasões estão relacionadas aos cursos de Engenharia Metalúrgica e as disciplinas obrigatórias são as que mais causam reprovações entre os alunos. O quadro 1 (um) apresenta o ranqueamento das disciplinas que possuem maiores reprovações.

**Tabela 1** - Ranqueamento das disciplinas com maiores reprovações

<b>Código disciplina</b>	<b>Nome disciplina</b>	<b>Casos</b>
GMA04043	Cálculo diferencial e integral Aplicado I	761 (14%)
GFI05100	Física geral e experimental XIII	674 (13%)
GAN06118	Álgebra linear aplicada	515 (10%)
GGM02055	Introdução à geometria descritiva	418 (8%)
TCC03060	Introdução à informática	350 (7%)
TCC03063	Programação de computadores III	234 (4%)
GMA06071	Equações diferenciais aplicadas	206 (4%)
GMA06074	Cálculo diferencial e integral aplicado II	171 (3%)
GMA04004	Cálculo diferencial e integral IV	156 (3%)
GFI05102	Física geral e experimental XX	155 (3%)
GFI05101	Física geral e experimental XIX	151 (3%)

Fonte: Elaborado pelos autores com base em SOUZA ( 2008)

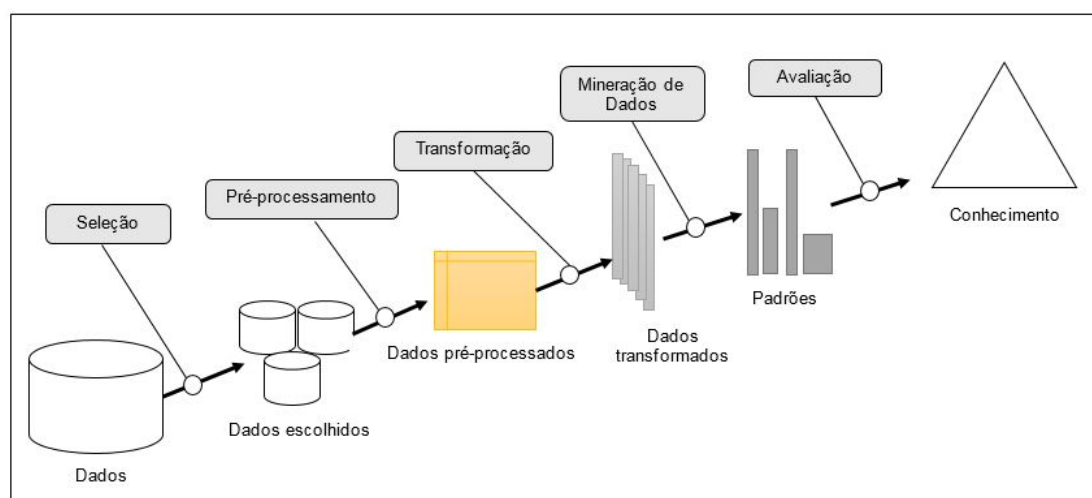
O INEP (2019a) traçou perfil dos ingressantes do ano de 2018 nas Redes Federais que mudaram de Unidade Federativa para estudar fora de seu local de residência. Do total de 309.266 ingressantes, 10,4% haviam desistido no primeiro ano de curso e 4,1% estavam com a matrícula trancada. Já no que tange ao contexto do Estado do Tocantins, que é objeto de estudo deste trabalho, 830 alunos foram oriundos de outros Estados (INEP, 2019a). Nota-se a necessidade de maior atenção para este grupo específico de alunos, pois a evasão escolar destes alunos pode estar relacionada a vários aspectos, como condições emocionais, educacionais ou ambientais. Nesse sentido, Souza (2008) assevera que a evasão nem sempre está relacionada ao baixo desempenho acadêmico e, por isso, é necessário investigar outros aspectos. Em suma, as IES, a fim de encontrar soluções para o problema de evasão escolar, devem analisar dados de seus acadêmicos, identificar situações eminentes de abandono e propor ações para minimizar a evasão escolar.

## **Metodologia**

Pela própria natureza dos dados acadêmicos da Unitins, a natureza da pesquisa é quantitativa e envolveu processo de coleta, análise e interpretação dos dados no intuito de descobrir relações entre as variáveis e características de causalidade ao fenômeno de evasão escolar. Como fundamentação teórica, foi utilizada a pesquisa bibliográfica a partir de materiais já elaborados e previamente publicados em livros e artigos científicos. Para o domínio dos dados foi aplicada a técnica de Mineração de Dados Educacionais (MDE) com a utilização de métodos de AM. Para isso foram utilizadas as ferramentas Spyder, linguagem de programação Python e os frameworks Scikit-learn e Pandas. Segundo (RAMESH e colab., 2013), a MDE auxilia na identificação de padrões para a tomada de decisão, cujo o processo é feito por meio da coleta dos dados, análise da informação e a geração do conhecimento. Já o AM fornece a base técnica para a MDE, que transforma dados brutos em informações de mais fácil compreensão, como previsões, correlações e relações de causalidade, o que, no processo de análise, auxilia na compreensão e explicação de fenômenos. Fayyad et al (1996) propuseram um processo denominado Descoberta de Conhecimento em Banco de Dados, do inglês Knowledge Discovery in Databases (KDD), com as seguintes

etapas: seleção, pré-processamento, transformação, mineração de dados e avaliação. As três primeiras etapas (seleção, pré-processamento e transformação) visam escolher, tratar, corrigir, normalizar e enriquecer os dados que serão processados para gerar conhecimento. Com os dados prontos para serem processados, a etapa de mineração de dados faz uso de algoritmos que extraem padrões dos dados e os padrões descobertos devem ser interpretados em uma fase de avaliação. A figura 1 (um) apresenta as principais etapas do KDD.

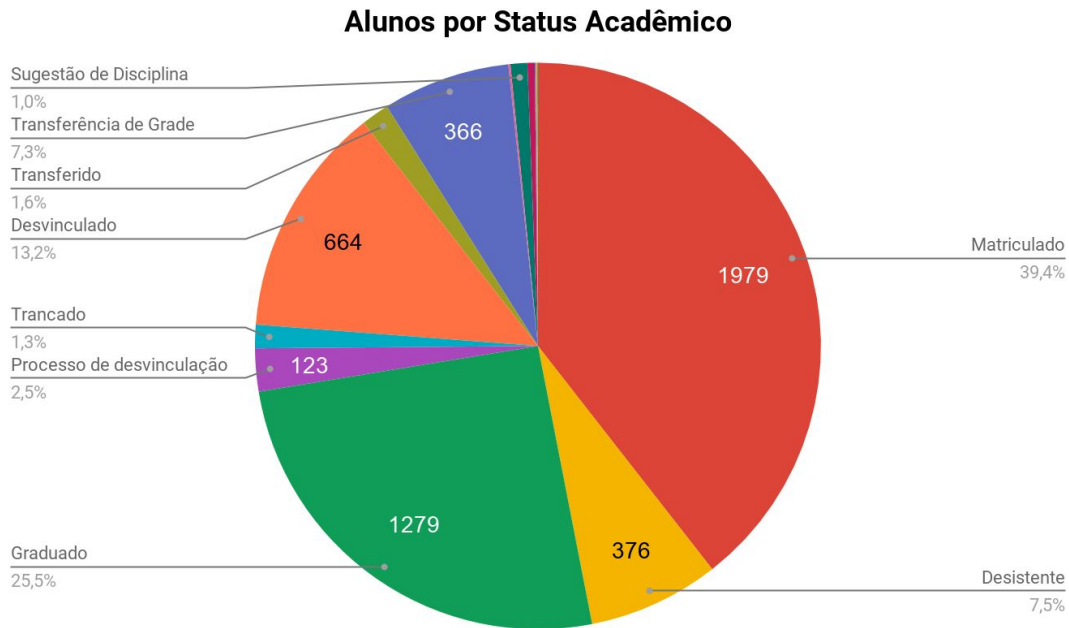
**Figura 1-** Principais etapas do processo de KDD.



Fonte: Elaborado pelos autores com base em Fayyad et. al. (1996)

Compreensão dos dados Para obtenção do acesso aos dados, esta pesquisa foi submetida e aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos (CEP), sob o número 29598820.3.0000.8023. Para a medição e treinamento dos classificadores foi utilizada a base de dados acadêmica dos 12 (doze) cursos presenciais da Unitins. Os dados estavam distribuídos em 18 (dezoito) variáveis de três tabelas (“cadastro de alunos”, “matrizes curriculares” e “histórico acadêmico”): RA, idade, sexo, estado civil, cor/raça, estado natal, naturalidade, UF, cidade, tipo de instituição, ano de formação do 2º grau, semestre de ingresso, tipo do ingresso, código da matriz curricular, último semestre cursado, câmpus, curso e turno. A figura 2 (dois) apresenta o quantitativo de alunos cadastrados na tabela de “cadastro de aluno”. Os acadêmicos estão classificados por status, que é a situação de vínculo em que eles se encontram.

**Figura 2 -** Quantitativo de alunos por tipo de status acadêmico



Fonte: Elaborado pelos autores com base em UNITINS (2020)

Posteriormente, foi feita uma seleção de dados a partir dos status acadêmicos mais relevantes para esta pesquisa: Graduado, Desvinculado e Desistente. Assim, do total de 5.018 (cinco mil e dezoito), foram selecionados 2.319 (dois mil e trezentos e dezenove) alunos. Um novo status, denominado Evadido, foi criado a partir da junção dos status Desistente e Desvinculado e o status graduado passou a se chamar Formado, para não gerar quaisquer ambiguidades com o status dos dados originais. Em seguida, na etapa de pré-processamento foi realizado preenchimento dos valores para os atributos que não estavam preenchidos. Para o atributo idade foi definida a média de idade da base de dados, os demais atributos foram registrados como “não informado”. Os dados não implícitos foram extraídos a partir do histórico acadêmico e da matriz curricular, na busca da identificação dos totalizadores, como: disciplinas da matriz, disciplinas cursadas, disciplinas aprovadas e disciplinas reprovadas por médias e faltas. A etapa de mineração de dados será abordada com detalhes na próxima seção.

### **Métodos de Aprendizado de Máquina**

O algoritmo utilizado como método de AM na etapa de mineração de dados foi da Floresta Aleatória (Random Forests). Segundo BREIMAN (2001), as melhorias

significativas na precisão da classificação resultaram o crescimento de um conjunto de árvores, que votam na classe mais popular, ficando conhecida como Florestas Aleatórias. De acordo com MONTAÑO (2016), esse método é mais adequado para bases de dados que possuem diversas variações e requer poucos ajustes de parâmetros, caso da base de dados utilizada neste trabalho. Para o treinamento do algoritmo, a base de dados foi particionada em dois conjuntos aleatórios, um com 70% dos dados a serem usados no treinamento, proporcional de acordo com as classes, e os demais 30% foram reservados para validação do treinamento. A validação cruzada (cross validation), que é uma técnica para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados, foi utilizada nas etapas de treinamento e avaliação. Na validação cruzada é feito particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, um elemento do conjunto é separado para testes (não participa do treinamento) e os demais elementos são utilizados no treinamento. Esse processo foi repetido k vezes alternando os elementos que foram utilizados para o grupo treinamento e para teste. A métrica de avaliação utilizada foi resultante das médias aritméticas dos k testes realizados (SCHAFFER, 1993). Os algoritmos de aprendizado de máquina, em geral, possuem parâmetros que podem ser ajustados para um melhor desempenho, segundo alguma métrica. Na etapa de validação cruzada, foi utilizado o algoritmo chamado Grid Search que cria uma grade multidimensional com a combinação dos parâmetros a serem avaliados. Para o algoritmo de Florestas Aleatórias, dois parâmetros foram testados: o critério de entropia e a quantidade de estimadores (árvores da floresta). Os critérios de entropia foram o padrão (entropy) e o critério de Gini. Enquanto que os valores de estimadores foram testados num intervalo entre 10 (dez) e 120 (cento e vinte). Estes valores do intervalo foram obtidos de forma empírica e em geral são dependentes do problema avaliado. Durante a etapa de validação cruzada, buscou-se obter a combinação de valores que maximizou o valor da acurácia, métrica que será explicada a seguir.

### **Métricas de avaliação da solução proposta**

Em geral, sistemas de aprendizado de máquina são avaliados por meio de métricas pontuais, aplicadas tanto na base de dados de treinamento quanto de teste.

De forma a subsidiar a discussão das métricas a serem adotadas para a solução proposta, a seguir, são explicadas duas métricas de modelos de predição considerando um problema com apenas duas classes (identificadas neste trabalho como positivo e negativo), que poderão ser utilizadas para avaliar e analisar os resultados gerados através da etapa de Mineração de Dados. Essas métricas são: Matriz de Confusão, que permite identificar a frequência das classes no modelo avaliado; A Receiver Operating Characteristic (ROC) e Área sob a Curva ROC ou Area Under the Curve (ROC-AUC), que serve para avaliar a qualidade do modelo.

229 Revista Humanidades e Inovação v.7, n.8 - 2020

A primeira maneira de se observar o desempenho de um classificador binário é através de uma Matriz de Confusão. Esta consiste de uma matriz quadrada onde são dispostas as previsões (linhas) e os valores verdadeiros (colunas). Ela é utilizada para mostrar a quantidade de acertos e erros de maneira que seja possível verificar a quantidade de amostras confundidas pelo sistema. Para tanto, é preciso definir o significado dos seguintes termos: Verdadeiro Positivo (VP), Falso Positivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN) (RASCHKA, 2015). Para o problema proposto neste trabalho, Verdadeiro Positivo (VP) é a quantidade de alunos que evadiram e o modelo identificou como evadido. O Verdadeiro Negativo (VN) é a quantidade de alunos que não evadiram e o modelo reportou como não evadido. Falso Positivo (FP) é a quantidade de alunos que não evadiram, mas foram erroneamente classificados como tal. Falso Negativo (FN) é a quantidade de alunos que evadiram, mas foram classificados como não evadido. Neste sentido, minimizar o FN é importante uma vez que, ao não identificar o aluno que evadiu, ações que poderiam evitar a evasão nestes casos podem deixar de ser tomadas. Com os valores definidos de VP, FP, VN e FN é construída a matriz de confusão, a qual é exibida na Figura 3.

**Figura 3-** Matriz de confusão com as duas classes (Evadido e Não Evadido)

Condição Atual  Predição	<b>Evadido (P)</b>	<b>Não Evadido (N)</b>
	<b>Aluno Evadido (P)</b>	VP (Verdadeiro Positivo)
<b>Aluno não Evadido (N)</b>	FN (evadido, mas não é extraído)	VN (Verdadeiro Negativo)

Fonte: Elaborado pelos autores com base em (RASCHKA, 2015)

A métrica acurácia ou exatidão (AC) diz o quanto o modelo acertou dentro de seu desempenho geral. A acurácia é obtida tomando-se a soma das previsões corretas, e dividindo-a pelo número total de previsões, respectivamente, conforme mostra a equação (1):

$$AC = \frac{VP+VN}{FP+FN+VP+VN} \quad (1)$$

Devido à bases de dados desbalanceadas, onde existem diferenças muito grandes entre o número de amostras de cada classe, em geral também utiliza-se outras métricas tais como Precisão e Revogação, F1-score curva ROC. Para dar suporte a estas métricas, é necessário estabelecer a Taxa de Falsos Positivos (FPR) e Taxa de Verdadeiro Positivo (TPR), que são definidas pelas equações (2) e (3), respectivamente. Estas taxas fornecem informações importantes, para as amostras positivas que foram identificadas corretamente no conjunto de total das amostras positivas.

$$FPR = \frac{FP}{N} = \frac{FP}{FP+VN} \quad (2)$$

$$TRP = \frac{VP}{P} = \frac{VP}{FN+VP} \quad (3)$$

Posto isto, define-se a Precisão (PRE) como a razão entre o número de verdadeiros positivos (VP) e a soma de Verdadeiro Positivos (VP) e Falso Positivos (FP), definida na equação (4). A métrica PRE serve para medir se daqueles que foram classificados como positivos, efetivamente estão corretos. Já a revogação ou *recall* (REC), apresentada na equação (5), é quão completo os resultados estão,



onde o REC de fato é o sinônimo de TPR. O REC pode ser realizado em situações que os FN são considerados mais prejudiciais do que os FP.

$$PRE = \frac{VP}{VP+FP} \quad (4)$$

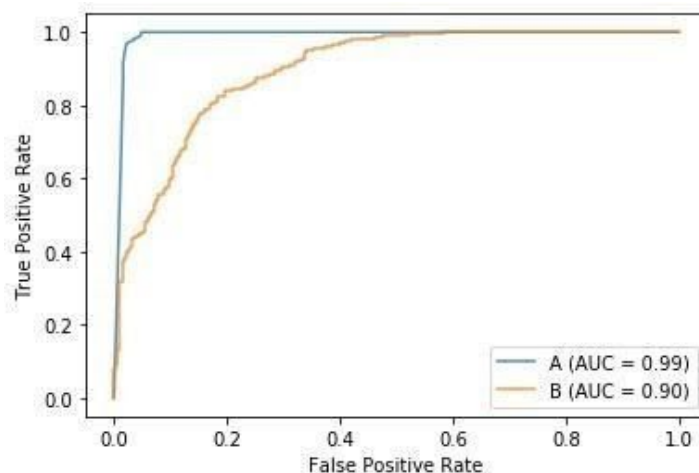
$$REC = TPR = \frac{VP}{P} = \frac{VP}{FN+VP} \quad (5)$$

Para combinar as duas informações PRE e REC em uma medida só, há uma métrica chamada F1-Score, definida na equação (6). A F1-Score é considerada uma média harmônica entre duas medidas PRE e REC, permitindo que seja possível a identificação da qualidade geral do modelo, através de um único valor. Portanto, valores altos de F1 implicam em uma qualidade maior do modelo avaliado.

$$F1 = 2 \frac{PRE \times REC}{PRE+REC} \quad (6)$$

De acordo com (MARTÍNEZ-CAMBLOR e PARDO-FERNÁNDEZ, 2019), a curva de ROC é uma ferramenta utilizada na estatística para estudar a qualidade de um processo de classificação em termos de sensibilidade e especificidade. Quando a curva padrão estiver com os valores maiores no marcador de diagnóstico, significa que existe maior probabilidade de se manter as características estudadas. Um gráfico ROC é a relação entre TPR e FPR. A figura 4 apresenta a curva ROC gerada para dois classificadores A e B. No gráfico da figura 11, o classificador A é interpretado como sendo melhor do que B, segundo a curva ROC. Esta medida pode ser simplificada através do cálculo da área sob a curva, chamado de ROC-AUC, que naturalmente varia de 0 (zero) a 1 (um) (RASCHKA, 2015).

**Figura 4 - Curva AUC (A e B)**



Fonte: Elaborado pelos autores com base em (MELO, 2013)

Portanto, a avaliação do modelo de classificação pode ser feita por meio de diferentes métricas de acordo com as características dos dados. É importante que se leve em consideração, durante a avaliação de modelos de classificação, fatores como proporção de dados de cada classe e o objetivo da precisão.

## Resultados

Com base nas métricas de Aprendizado de Máquina, apresenta-se os resultados que foram obtidos para a avaliação do modelo. As métricas utilizadas para este estudo foram: Acurácia, Precisão, Revogação, F1-Score e ROC-AUC.

A tabela 2 (dois) expõe os resultados obtidos no conjunto treinamento, que utilizou 70% da base de dados, com o algoritmo Floresta Aleatória pela técnica de validação cruzada. Como resultado, destaca-se acurácia média de 97,4%, entre os subconjuntos gerados pelo processo de validação cruzada.

**Tabela 2 - Resultados dos treinamentos**

Métricas de Aprendizado de Máquina	Resultado
Acurácia	97,4
Precisão	96,2
Revogação	99,2
F1	97,6

ROC-AUC	97,1
---------	------

Fonte: Elaborado pelos autores (2020)

Como métrica de avaliação dos resultados, a matriz de confusão foi calculada e está apresentada na tabela 3 (três). Os dados contidos na tabela 3 (três) representam a média gerada pela validação cruzada e apresenta a porcentagem de acertos da metodologia adotada neste trabalho em relação à sua capacidade de classificar alunos evadidos e não evadidos. Destaca-se que 88,8% dos alunos que não evadiram, chamado de Verdadeiro Negativo (VN), e que 69,3% dos acadêmicos evadidos, chamado de Verdadeiro positivo (VP), foram classificados corretamente.

**Tabela 3** - Matriz de Confusão gerada a partir do treinamento

Matriz de Confusão - Validação Cruzada	
VP (69,3%)	FP (3,5%)
FN (0,7%)	VN (88,8%)

Fonte: Elaborado pelos autores (2020)

Em seguida, o algoritmo de Floresta Aleatória, com seus parâmetros otimizados pelo processo de *Grid Search*, foi treinado nos 70% da base de dados e testado utilizando os demais 30%, que não foram apresentados ao classificador, em qualquer momento, durante o treinamento. A Tabela 4 (quatro) traz os resultados alcançados.

**Tabela 4** - Resultado dos classificadores dos grupos de 70% de treinamento e 30% para teste

Métricas de Aprendizado de Máquina	Resultado
Acurácia	97,5

Precisão	95,7
Revogação	100
F1	97,8
ROC-AUC	97,2

Fonte: Elaborado pelos autores (2020)

Percebe-se que os resultados dos dados de validação (30% da base de dados), obtêm-se os resultados semelhantes com os alcançados durante o treinamento (70% dos dados). Este fato é um indicador de que o modelo se adaptou (aprendeu) de tal forma que conseguiu generalizar o conhecimento para novos dados (objetivo principal de qualquer preditor).

A matriz de confusão da tabela 5 (cinco) apresenta a média gerada pela validação cruzada e expressa a quantidade de acertos em classificar a situação do acadêmico: evadido ou não evadido. Nesse sentido, observa-se o bom desempenho dos métodos utilizados neste trabalho na obtenção de VP (alunos evadidos) e de VN (alunos não evadidos).

**Tabela 5** - Resultado da Matriz de Confusão do grupo 70% de treinamento de 30% para teste.

Matriz de Confusão - Grupos de 70 e 30%	
VP (295)	FP (17)
FN (0)	VN (384)

Fonte: Elaborado pelos autores (2020)

Observando os resultados apresentados na Tabela 5 (cinco), percebe-se que nos 30% aleatoriamente selecionados, e conseqüentemente não utilizados durante a

fase de treinamento, a quantidade de Falsos Negativos (FN) é zero. Observando a matriz de confusão exibida na Tabela 3 (três), que remete aos dados de treinamento, percebe-se que este valor é menor que 1%. Ou seja, nos 70% dos dados, a validação cruzada separou em  $k=10$  grupos, e o algoritmo foi treinado por 10 (dez) vezes seguidas, sempre utilizando-se 9 (nove) subgrupos para treinamento e 1 (um) para testes. Ainda neste cenário, o valor médio dos FN foi menor também próxima de zero.

É interessante ressaltar que o Falso Negativo representa o caso em que o aluno evadiu, mas foi classificado como não evadido. Este caso é o mais sensível para a aplicação considerada. Perceba que, se o sistema classificar um aluno provável de evadir como se não evadisse, os gestores do ensino não poderiam também considerar a situação deste aluno para criar elementos para tentar alcançá-lo, antes que este decida por evadir. Observando os Falsos Positivos, errar neste aspecto significaria errar por excesso, pois eles também seriam considerados para elaboração de políticas para lidar com o problema. Ainda analisando a Tabela 5 (cinco), percebe-se que 17 (dezesete) amostras se encaixam nesta situação. Isto é apenas uma proporção de 0,02 (dois centésimos), quando comparado com o total de amostras considerada na base de dados de validação. O excesso é mínimo e, portanto, suportável pelo poder informativo do experimento.

### **Considerações Finais**

A evasão escolar é um problema que afeta todos os níveis de ensino. Os aspectos relacionados a esse fenômeno são diversos e complexos e nem sempre estão associados ao âmbito acadêmico. O uso de ferramentas de inteligência computacional tais como técnicas de AM, tem contribuído bastante para estudos sobre Evasão Escolar ao extrair, de bases de dados, informações e conhecimentos que podem subsidiar o processo de tomada de decisão das IES.

Este trabalho validou, por meio de um estudo de caso, a hipótese de que as técnicas AM podem ser usadas de maneira satisfatória em MDE, o que foi verificado por experimentos utilizando a base de dados acadêmica da Unitins. Os resultados alcançados neste trabalho indicam que os métodos de MDE podem contribuir para o desenvolvimento de um classificador de confiança elevada, ou seja, capaz de

automatizar a tarefa de classificar alunos em estado de evasão. Uma vez feita esta separação, é possível conduzir um estudo mais focado no grupo identificado e tentar encontrar padrões que levaram à evasão.

Sabe-se que os setores acadêmicos que lidam com diversos processos e, em geral, um grupo pequeno de trabalho é alocado na tarefa de tentar entender as razões da evasão nas instituições e propor soluções. O experimento computacional reportado nesta pesquisa dá fortes indicativos de que é possível aproximar um modelo de evasão a partir dos registros acadêmicos.

Como trabalhos futuros, pretende-se conduzir uma investigação sobre a capacidade de um sistema de aprendizado de máquina de agir como preditor em dados parciais. Visto que, ao final de cada semestre letivo, o sistema poderia propor uma predição de evasão e reportar este percentual aos coordenadores de curso ou para a gestão acadêmica. Uma vez de posse destes dados, poderiam tomar decisões fundamentadas e mais acuradas para alcançar o aluno provável de evadir, antes do acontecimento em si. Por fim, é interessante ressaltar que o fator humano será cada vez mais necessário, uma vez que a máquina será capaz de indicar a porcentagem de previsão de evasão, mas o entendimento do que poderia levar o aluno a evadir terá que ser interpretado por outro ser humano.

## Referências

ADACHI, Ana Amélia Chaves Teixeira. **Evasão de estudantes de cursos de graduação da USP – Ingressantes nos anos de 2002, 2003 e 2004**. 2017. 294p. Tese. (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo-SP, 2017.

BAGGI, Cristiane Aparecida Dos Santos e DOS SANTOS BAGGI; Cristiane Aparecida e LOPES, Doraci Alves. **Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica**. Avaliação: Revista da Avaliação da Educação Superior (Campinas). [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s1414-40772011000200007>>. , 2011

BRASIL, Lei nº 9.394, de 20 de dezembro de 1996a. **Estabelece as diretrizes e bases da educação nacional**. Brasília, DF, dez 1996.

BRASIL. Ministério da Educação. Secretaria de Educação Superior (1996b). Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. **Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas**. Acesso em: 11/04/2020. Disponível em: [http://www.andifes.org.br/wp-content/files\\_flutter/Diplomacao\\_Retencao\\_Evasao\\_Graduacao\\_em\\_IES\\_Publicas-1996.pdf](http://www.andifes.org.br/wp-content/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf).

BREIMAN.L. **Random forests**. Machine Learning, 45:5–32, 2001

CARVALHO, Alessandro Pires (2017). **Fatores institucionais associados à evasão na educação superior**. 2017. 90 f. Dissertação. (Mestrado em Administração)-Programa de Pós-Graduação em Administração, Universidade Federal de Goiás, Goiânia.

COLPANI, Rogério. **Mineração de Dados Educacionais: um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar**. Informática na educação: teoria & prática. [S.l.: s.n.]. Disponível em: <<http://dx.doi.org/10.22456/1982-1654.87880>>. , 2019

CUNHA, E.R, MOROSINI, M.C.(2013). **Evasão na Educação Superior: Uma temática em Discussão**. Disponível em:<https://paginas.uepa.br/seer/index.php/cocar/article/view/283>. Acesso em: 14 de dez 2019.

FAYYAD, U. M., Piatetsky Shapiro, G., Smyth, P. & Uthurusamy, R. “**Advances in Knowledge Discovery and Data Mining**” 1996, AAAIPress, The Mit Press.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. (1996); **From Data Mining to Knowledge Discovery in Databases**. Disponível em: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>. Acesso em: 28 de fev. 2020.

FILHO, Roberto Leal Lobo e. Silva e colab. **A evasão no ensino superior brasileiro. Cadernos de Pesquisa**. [S.l.: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s0100-15742007000300007>>. , 2007

GAIOSO, N. P. L. **O fenômeno da evasão escolar na educação superior no Brasil**. 2005. 75 f. Dissertação (Mestrado em Educação) Programa de Pós-Graduação em Educação da Universidade Católica de Brasília, Brasília, DF, 2005.

GUIMARÃES, Orlineya Maciel; MARTINS, Eliana Canteiro Bolorino; LIMA, Maria Jose de Oliveira (2020). **A EVASÃO NO ENSINO SUPERIOR: A UNESP CÂMPUS DE FRANCA - PERÍODO DE 2013-2018. CAMINE: Caminhos da Educação = Camine: Ways of Education**, Franca, v. 11, n. 2, p. 136-161, mar. 2020. ISSN 2175-4217. Disponível em: <<https://periodicos.franca.unesp.br/index.php/caminhos/article/view/3013>>. Acesso em: 15 abr 2020.

GOMES, E. C.; SOARES, D. B.; DESIDÉRIO, S. N.; ROCHA, A. S. DA R. S. DA. (2019). **EVASÃO NO CURSO DE LICENCIATURA EM FÍSICA DA UNIVERSIDADE FEDERAL DO TOCANTINS: diagnóstico e primeiros resultados de um projeto de intervenção**. Revista Observatório, v. 5, n. 5, p. 482-508, 1 ago.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019a). **Censo da Educação Superior 2018**. Divulgação dos Resultados. Brasília- DF- 19 de Setembro de 2019.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019b). **Sem desistências, número de graduados poderia dobrar no Brasil**. [http://inep.gov.br/artigo/-/asset\\_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206](http://inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206). Acesso em 08 out 2019.

KIRA, L. P. (1998) **A evasão no ensino superior: o caso do curso de pedagogia da Universidade Estadual de Maringá (1992-1996)**. Dissertação (Mestrado em Educação), Universidade Metodista de Piracicaba, 106 p.

MARTÍNEZ-CAMBLOR, Pablo e PARDO-FERNÁNDEZ, Juan C. **Parametric estimates for the receiver operating characteristic curve generalization for non-monotone relationships**. Statistical methods in medical research, v. 28, n. 7, p. 2032–2048, Jul 2019.

MELO, Francisco. **Area under the ROC Curve**. Encyclopedia of Systems Biology. [S.l.]: Springer, New York, NY, 2013. p. 38–39. . Acesso em: 31 out 2019.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto (2003). **Conceitos sobre Aprendizado de Máquina**. Disponível em: <<http://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>>. Acesso em: 9 dez 2019.

MONTAÑO. RAZER. **Aplicação de Técnicas de Aprendizado de Máquina na Mensuração Florestal**. Disponível em: <https://www.acervodigital.ufpr.br/bitstream/handle/1884/45346/R%20-%20T%20-%20RAZER%20ANTHOM%20NIZER%20ROJAS%20MONTANO.pdf?sequence=1&isAllowed=y>. Acesso em: 16 Mar 2020.

RAMESH, V. e PARKAVI, P. e RAMAR, K. **Predicting Student Performance: A Statistical and Data Mining Approach**. International Journal of Computer Applications. [S.l.: s.n.]. Disponível em: <<http://dx.doi.org/10.5120/10489-5242>>. , 2013

RASCHKA, Sebastian. Python Machine Learning. [S.l.]: Packt Publishing Ltd, 2015.

SCHAFFER. C. **Selecting a classification method by cross-validation**. Disponível em: <https://link.springer.com/content/pdf/10.1007/BF00993106.pdf>. Acesso em: 17 Mar 2020.

SABBATINI, M. (2015). **Concepções e estratégias da aprendizagem participativa na educação a distância (EAD): contribuição das práticas dialógicas e comunicacionais para a autonomia discente**. Revista Observatório, v. 1, n. 3, p. 80-99, 26 dez.

SANTOS JUNIOR, J. DA S.; REAL, G. C. M (2019). **Fator institucional para a evasão na educação superior**. Revista Internacional de Educação Superior, v. 6, p. e020037, 27 dez.

SOUZA, Solange Lima de (2008). **Evasão no Ensino Superior: Um estudo utilizando a mineração de dados como ferramenta de gestão do conhecimento em um banco de dados referente à graduação de Engenharia**. Disponível em: <http://livros01.livrosgratis.com.br/cp064905.pdf>. Acesso: 11 abr 2020.

SOUZA, Thays Santos (2017). **Estudo sobre a evasão em cursos de graduação presenciais na Universidade Federal de Goiás – UFG**. 2017. 214 f. Dissertação. (Mestrado Profissional em Gestão Organizacional)-Programa de Pós-Graduação em Gestão Organizacional, Universidade Federal de Goiás, Catalão.

UNITINS (2020). Universidade Estadual do Tocantins. **Graduação**. Disponível em: <https://www.unitins.br/nportal/graduacao>. Acesso em: 11 abr 2020.



**ARTIGO SUBMETIDO PARA A REVISTA DADOS - REVISTA DE  
CIÊNCIAS SOCIAIS**

ISSN: 0011-5258

**Sistema Inteligente para Previsão de Evasão Escolar em um Contexto Universitário**

**Resumo:** O fenômeno da Evasão Escolar é um dos problemas existentes no sistema de ensino brasileiro e que atinge seus mais variados níveis. Especificamente ao Ensino Superior, este problema é considerado de âmbito internacional, tornando-o ainda mais impactante. Em geral, as universidades sabem as taxas gerais de evasão por curso, e ainda que saibam exatamente os alunos evadidos, prever quais poderão ainda evadir é uma tarefa complexa. Neste sentido, este trabalho apresenta uma metodologia de solução que utiliza-se de modelos computacionais inteligentes, especificamente o aprendizado de máquina para aprender os padrões dos alunos evadidos e não-evadidos. Isto significa, que desde o primeiro período a solução permite inferir a probabilidade do aluno evadir. Um estudo de caso foi realizado com os dados acadêmicos da Universidade Estadual do Tocantins (Unitins). Os resultados alcançados indicaram que a metodologia adotada neste trabalho permite classificar os alunos propensos à evasão, com elevado grau de confiança.

**Palavras-chave:** Evasão Escolar; Mineração de Dados Educacionais; Aprendizagem de Máquina; Inovação em Gestão Educacional

Autores: Leandra Cristina Cavina Piovesan Soares, Robson Aparecido Ronzani, Rafael Lima de Carvalho.

## 6 Sistema Inteligente para Previsão de Evasão Escolar em um Contexto Universitário

**Resumo:** O fenômeno da Evasão Escolar é um dos problemas existentes no sistema de ensino brasileiro e que atinge seus mais variados níveis. Especificamente ao Ensino Superior, este problema é considerado de âmbito internacional, tornando-o ainda mais impactante. Em geral, as universidades sabem as taxas gerais de evasão por curso, e ainda que saibam exatamente os alunos evadidos, prever quais poderão ainda evadir é uma tarefa complexa. Neste sentido, este trabalho apresenta uma metodologia de solução que utiliza-se de modelos computacionais inteligentes, especificamente o aprendizado de máquina para aprender os padrões dos alunos evadidos e não-evadidos. Isto significa, que desde o primeiro período a solução permite inferir a probabilidade do aluno evadir. Um estudo de caso foi realizado com os dados acadêmicos da Universidade Estadual do Tocantins (Unitins). Os resultados alcançados indicaram que a metodologia adotada neste trabalho permite classificar os alunos propensos à evasão, com elevado grau de confiança.

**Palavras-chave:** Evasão Escolar; Mineração de Dados Educacionais; Aprendizagem de Máquina; Inovação em Gestão Educacional.

**Abstract:** The school dropout phenomenon is one of the problems that exist in the Brazilian education system and that reaches its most varied levels. Specifically to Higher Education, this problem is considered international, making it even more impactful. In general, universities know the general dropout rates per course, and even though they know exactly which dropout students are, predicting which dropout students are likely to drop out is a complex task. In this sense, this work presents a solution methodology that uses intelligent computational models, specifically machine learning to learn the patterns of escaped and non-escaped students. This means that from the first period the solution allows inferring the probability of the student to evade. A case study was carried out with academic data from the State University of Tocantins (Unitins). The results achieved indicated that the methodology adopted in this work allows classifying students prone to dropout, with a high degree of confidence.

**Keywords:** School dropout; Educational Data Mining; Machine Learning; Innovation in Educational Management.

### INTRODUÇÃO

A temática sobre o fenômeno da evasão escolar faz parte de debates e reflexões entre o meio acadêmico (SABBATINI, 2015; ADACHI, 2017; CARVALHO, 2017; SOUZA, 2017; GUIMARÃES et al, 2019; GOMES et al., 2019; SANTOS JUNIOR; REAL, 2019). Estudos como (TINTO, 1975; BRASIL, 1996a) auxiliam no entendimento dos motivos pelos quais os estudantes deixam o seu curso, reforçando a necessidade de uma maior investigação visando a identificação das causas e motivos associados. Associa-se de forma generalista, que a evasão está relacionada ao curso, mas não necessariamente este seja o problema. Nota-se que a evasão escolar é composta por diversas variáveis que se interagem e se conflitam em torno da

problemática. Estudos como de (SOUZA, 2008; BAGGI e colab., 2011; BARLEM [et al. 2012](#)) relatam as possíveis motivações que podem levar o aluno a abandonar o seu curso.

Para SOUZA (2008) a motivação da evasão não está restrita apenas no âmbito acadêmico, a causalidade pode advir de aspectos pessoais, sociais e ambientais. Segundo (BAGGI e colab., 2011) o abandono escolar requer uma análise histórica, a realidade dos níveis anteriores de ensino podem influenciar no abandono de um curso superior. No entendimento dos autores (BARLEM e colab., 2012) vários motivos levam o estudante a evadir-se de um curso, sendo ele pelo próprio universo discente, como imaturidade, desconhecimento do curso, dificuldades de adaptação, problemas financeiros e/ou familiares. Já no que tange ao ambiente acadêmico, pode ser relacionado a insatisfação com o sistema de ensino ou, até mesmo, insatisfação com a profissão escolhida. Nesse contexto, observa-se que este fenômeno merece uma atenção especial, dado a sua complexibilidade e representatividade na construção de um planejamento de políticas públicas da educação.

Neste sentido alguns autores trazem estudos como de (KIRA, 1998; GAIOSO, 2005; BAGGI e colab., 2011) que a evasão escolar é definida pela interrupção do ciclo de estudos, em qualquer nível de ensino. Outras definições como da Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras (CEUPB) a evasão escolar é caracterizada pela saída definitiva do aluno de seu curso de origem, sem concluí-lo (BRASIL, 1996). Os autores (SCHARGEL e SMINK, 2002) indicam que a evasão escolar é dividida em categorias de causa: psicológicas, sociológicas, organizacionais, interacionais e as econômicas. Além disso, os autores argumentam que o combate para fenômeno será efetivo, somente quando houver uma abordagem sistêmica, já que o problema não é isolado.

Enquanto que em (SOARES e colab., 2020) mostrou-se a possibilidade de treinar um sistema aprender os padrões de alunos evadidos ou não, o presente trabalho reporta uma metodologia de um modelo por período. Esta metodologia permitiu a geração de um modelo inteligente por período escolar de maneira a tornar o sistema de inferência mais preciso e orientado. Assim, o objetivo do presente estudo é apresentar os resultados do modelo de inteligência artificial treinado para inferir a probabilidade do aluno evadir, por período de matrícula. O estudo de caso reportado fez uso dos dados acadêmicos da Universidade Estadual do Tocantins (UNITINS). Neste sentido, a solução aqui apresentada permitirá os gestores escolares tomarem decisões mais orientadas aos indivíduos mais propensos a evadir,

possibilitando a minimização do problema de evasão. Além do mais, este estudo mostra como tecnologias de ponta podem ser aplicadas de maneira a minimizar o problema da evasão escolar.

## **A EVASÃO ESCOLAR NO ENSINO SUPERIOR BRASILEIRO**

A redução da taxa de desistência e a estada por longo tempo dos alunos nas Instituições de Ensino Superior (IES) são uns dos grandes desafios para a Educação Brasileira (INEP, 2019). Percebe-se que estudos anteriores já haviam sido realizados, como no ano de 1996 com a criação da Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras (CEUPB) que contava com a participação das Instituições de Ensino Superior Públicas, Federais e Estaduais. Cujo objetivo era o estudo do desempenho das Instituições de Ensino Superiores Brasileiras. Um das metas desta comissão, estava na definição das ações a serem trabalhadas. O quadro 1 apresenta as ações definidas pela CEUPB.

### **Quadro 1**

#### **Objetivos da CEUPB**

<b>Item</b>	<b>Ação</b>
1	Definir o conceito de evasão, considerando suas dimensões concretas: evasão de curso, evasão da instituição e evasão do sistema de ensino superior;
2	Definir e aplicar a metodologia igualitária de coleta e tratamento dos dados;
3	Identificar as taxas de diplomação, retenção e evasão dos cursos de graduação das IES do país;
4	Apontar as causas internas e externas da evasão, considerando as especificidades dos cursos e das regiões do país;
5	Definir estratégias voltadas à redução dos índices de evasão nas universidades públicas brasileira.

Fonte: Elaborado pelos autores com base em (BRASIL, 1996a).

Após anos de estudos empreendidos pela Comissão Especial de Estudos sobre Evasão nas Universidades Públicas Brasileiras (CEUPB) os resultados foram sendo alcançados, como a

definição do tipo de evasão e dentro outros. Mas como parecer final, a comissão reforça a necessidade de um estudo mais bem aprofundado, pois considera que este fenômeno é complexo e igualitário para as Instituições de Ensino Superior (IES), nos quais provocam a necessidade de estudos e análises sobre tal tema, já que este fenômeno é influenciado por diversas variáveis (BRASIL, 1996a). O texto apresentado no relatório da Comissão, reforça essa necessidade:

O reconhecimento dos óbices que condicionaram este estudo corrobora a certeza de que o conhecimento mais completo e confiável do fenômeno só poderá ser alcançado através de um verdadeiro programa integrado de pesquisas que estabeleça os elos entre os níveis, identifique causas internas e externas, dando assim a necessária dimensão de totalidade característica de uma avaliação do sistema de ensino superior público do país. (BRASIL, 1996a)

Portanto, no Brasil a Educação Superior é avaliada anualmente pelos indicadores do Censo Superior, efetuado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Pelos números é possível ter a indicação das taxas de evasão escolar brasileira no ensino superior. Considerando os dados compilados do Censo de 2018, referente aos alunos ingressantes do ano de 2010, foi possível retratar a tabela 1, com seus respectivos quantitativos. A tabela 1 apresenta os quantitativos de alunos .

**Tabela 1**

Dados informativos dos ingressantes de 2010 com recorte em 2016

<b>Ano de Ingresso</b>	<b>% de Alunos Desistentes</b>	<b>% Alunos formados</b>	<b>% Alunos de permaneceram na graduação por mais de 6 anos</b>
2010	56,8 %	37,9%	5,3%

Fonte: Elaborado pelos autores com base em (INEP, 2019)

Neste aspecto, estudos atuais buscam dar ênfase não apenas pelas razões pelo qual o aluno evadiu, mas também buscam a compreensão por meio de ações preventivas estimulando-os a permanecer no sistema. Assim, para a Administração Pública, a evasão e a continuação por longo tempo dos estudos, são obstáculos que causam reflexos institucionais e sociais,

levando-se em conta os investimentos em recursos humanos e financeiros (PEREIRA et al., 2011; BAGGI e colab., 2011).

Segundo os autores (RIGO e colab., 2014) a evasão escolar está presente tanto nas universidades públicas e privadas e a sua consequência são as questões financeiras com a redução de número de alunos formados no ensino superior, causando impacto na cadeia produtiva do nosso país. LOBO (2012) indica que os estudos sobre a evasão, requer um modelo de política de governo orientada a qualidade do ensino. Com incentivos financeiros voltados para a pesquisa e estudos que possam identificar e combater com eficiência este fenômeno. Corroborando com (LOBO, 2012) a Associação de Mantenedoras de Ensino Superior (ABMES) reforça também sobre a necessidade da criação de políticas governamentais, com ações orientadas para a qualidade do ensino e uso dos recursos públicos e privados, sendo conduzidos para a promoção de processos e análises direcionadas na realização de atividades (HORTA, 2012).

Nos estudos de (MARTINHO, 2014) foi desenvolvido um sistema preditivo inteligente, para prever o risco de evasão dos alunos dos cursos Superiores de Tecnologia (CST) em Automação e Controle da Universidade Federal do Mato Grosso (IFMT). Os dados acadêmicos compreendidos para a análise foram dos anos de 2004-2 a 2011-2. As técnicas utilizadas são da Inteligência Artificial, especificamente Redes Neurais Artificiais (*ARTMAP-Fuzzy*) e os resultados indicaram altos índices de acertos do grupo de alunos propensos a evadir, atingindo um percentual de 95% e 100% de acurácia e uma média global de 95% de acertos.

Os dados do CENSO também podem ser fontes de pesquisas para este tipo de fenômeno. Estudos como de (COLPAN, 2019) utilizou os dados do CENSO de 2017 e como resultado descobriu que o Estado do Pará é o maior em índices de evasão no Brasil. Em sua pesquisa, foi diagnosticado que a média de evasão escolar no Estado é de 20%. Deste percentual, estão em uma escala de 38% a 65% dos alunos com idade acima da recomendada que é de 18 anos, previstas na Lei 9.394/96 (BRASIL, 1996b).

Assim sendo, este fenômeno no ensino superior é considerado um dos principais desafios a serem superados. Os trabalhos que busquem responder a estas indagações é essencial para o estabelecimento de um campo de discussão, auxiliando as Instituições de Ensino Superior nas

políticas educacionais, contribuindo assim, para uma educação superior de qualidade no para o nosso país. E propor soluções que busquem meios de identificar de forma proativa e acurada o grupo dos alunos propensos a evadir tornam-se cada vez mais importante para as IES. Portanto, o desenvolvimento de um Sistema inteligente permite que a IES se posicione de forma a antecipar os problemas que possam surgir futuramente, como a redução dos seus índices de evasão.

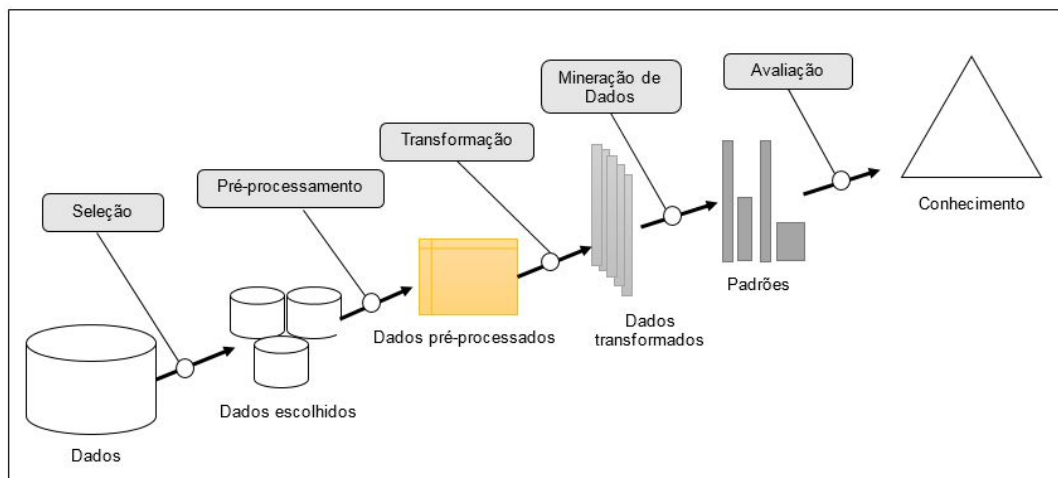
## **MATERIAIS E MÉTODOS**

A natureza da pesquisa adotada é a quantitativa, onde envolveu o processo de coleta, análise e interpretação dos dados acadêmicos com a intenção de descobrir relações entre as variáveis e características de causalidade sobre a evasão escolar. Como fundamentação teórica utilizou-se da pesquisa bibliográfica a partir de materiais já elaborados e previamente publicados em livros e artigos científicos. Para a compreensão dos dados aplicou-se o uso de ferramentas computacionais inteligentes tais como as técnicas de Mineração de Dados Educacionais (MDE) e os métodos de Aprendizado de Máquina (AM).

De acordo com os autores ([RIGO e colab., 2014](#)) com o aumento substancial do uso das tecnologias de informação e comunicação (TIC) fez com que aumentasse substancialmente as bases de dados, fazendo com que a capacidade de geração dos dados fosse maior do que a conhecimento dos pesquisadores e analistas. Este cenário também está presente no ambiente acadêmico, onde as instituições de ensino possuem uma gama de sistemas informatizados, como ambiente de educação a distância, sistemas acadêmicos e comunidades virtuais. A partir deste cenário, a busca pela exploração dos dados dos ambientes educacionais, por novas descobertas fez surgir uma nova área de pesquisa chamada de Mineração de Dados Acadêmicos (MDE) ou *Educational Data Mining*, (EDM).

Para tal, a pesquisa utilizou-se de ferramentas tecnológicas como: *Spyder*, linguagem de programação *Python*, *frameworks Scikit-learn* e *Pandas*. Para a extração do conhecimento na base de dados ou *Knowledge Discovery in Databases* (KDD), aplicou-se algoritmos de análises e descobertas de dados sob controles de eficiência computacional aceitável. Segundo FAYYAD et al (1996) a descoberta do conhecimento em banco de dados é constituída por diversas fases, sendo: a seleção, o pré-processamento, a transformação, a mineração de dados e avaliação. A figura 1 apresenta o detalhamento deste processo do KDD.

**Figura 1**  
**Etapas do processo de KDD**

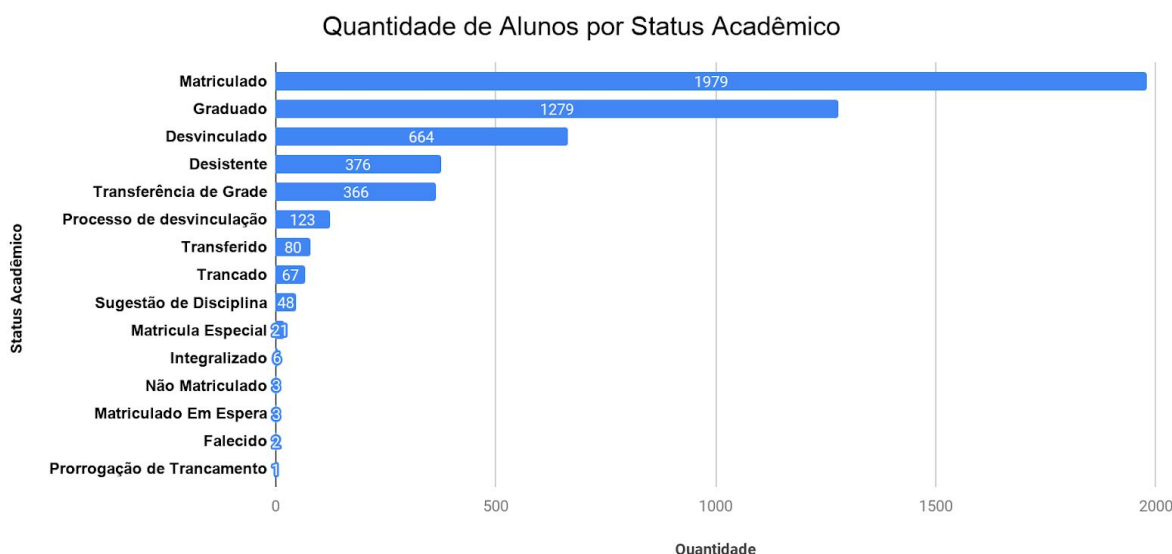


Fonte: Elaborado pelos autores com base em FAYYAD et. al. (1996)

Para a medição e treinamento dos classificadores, foram utilizados os 12 (doze) cursos presenciais da Universidade Estadual do Tocantins (Unitins). As tabelas utilizadas foram: Cadastro de alunos, Matrizes curriculares e Histórico acadêmico. Após análise das tabelas, foram selecionadas 18 (dezoito) variáveis, sendo: RA, idade, sexo, estado civil, cor/raça, estado natal, naturalidade, UF, cidade, tipo de instituição, ano de formação do 2º grau, semestre de ingresso, tipo do ingresso, código da matriz curricular, último semestre cursado, câmpus, curso e turno. A partir da extração dos dados da tabela “Cadastro de Alunos”, chegou-se nos quantitativos de alunos por status acadêmico. O status acadêmico é um dado importante, pois é através dele que serão identificados os alunos que farão parte dos testes e treinamentos dos classificadores. A figura 2 (dois) apresenta os quantitativos de alunos por status acadêmico. A seguir serão apresentados os status acadêmicos mais relevantes para a pesquisa.

**Figura 2**  
**Quantitativo de alunos por tipo de status acadêmico**





Fonte: Elaborado pelos autores (2020)

Posteriormente, houve a seleção de dados a partir dos *status* acadêmicos mais expressivo para a pesquisa: Graduado, Desvinculado e Desistente. Houve um filtro a partir do quantitativo de 5.018 (cinco mil e dezoito). Deste total, foram selecionados 2.319 (dois mil e trezentos e dezenove) alunos que atendiam os critérios da seleção. Um novo *status* chamado Evadido foi formado a partir da junção dos *status* Desistente e Desvinculado. Assim, o *status* graduado foi transferido a sua nomenclatura para Formado, para não gerar quaisquer ambiguidades com o *status* dos dados originais.

A próxima etapa é chamada de pré-processamento, onde foram preenchidos os valores para os atributos que estavam nulos. Os dados não informados do atributo idade foram substituídos pela média da idade, enquanto que os demais atributos nulos ou vazios foram registrados como “não informado”. Os dados não implícitos foram extraídos a partir do histórico acadêmico e da matriz curricular, como forma de buscar a identificação dos totalizadores, como: disciplinas da matriz, disciplinas cursadas, disciplinas aprovadas e disciplinas reprovadas por médias e faltas. Para melhor exemplificar essa etapa, a tabela 2 apresenta os quantitativos de dados por atributos de classe.

**Tabela 2**  
**Quantitativo de Atributos com valores nulos**

Qtde de atributos com valores nulos	Etapa de pré-processamento dos dados					
	Qtde de registros da Classe “Evadido”	% Classe Evadido	Qtde de registros da Classe “Graduado”	% Classe Graduado	Total das classes	% Total das classes
Nenhum	495	47,60%	730	57,08%	1225	52,85%
1	455	43,75%	407	31,82%	862	37,19%
2	72	6,92%	37	2,89%	109	4,70%
3	16	1,54%	103	8,05%	119	5,13%
4	0	0,00%	2	0,16%	2	0,09%
5	1	0,10%	0	0,00%	1	0,04%
6	1	0,10%	0	0,00%	1	0,04%
<b>Total de registros</b>	<b>1040</b>		<b>1279</b>		<b>2319</b>	

Fonte: Elaborado pelos autores (2020)

Nota-se na tabela 2 (dois), que o maior percentual equivale a coluna “nenhum”. Percebe-se que 52,85% dos atributos pertencentes a esta classe, estão com dados preenchidos, não sendo necessário o manuseio e/ou preenchimento de dados ausentes. A partir da tabela 2 (dois), foi construída a tabela 3 (três) que apresenta os atributos que possuem maiores quantitativos nulos.

**Tabela 3**  
**Atributos com maiores quantidades nulas**

Atributo	Descrição	Qtde de registros da Classe “Evadido”	Qtde de registros da Classe “Graduado”	Total das classes
CORRACA	Cor e raça	520	463	983
ANOULTIMAINST	Último ano cursado no	56	105	161

	ensino médio			
ULTIMO_SEMESTRE	Último semestre cursado	0	103	103
ESTADOCIVIL	Estado civil	67	24	91
<b>TOTAL</b>		<b>643</b>	<b>695</b>	<b>1338</b>

Fonte: Elaborado pelos autores (2020)

Observando-se os números do atributo CORRACA é o que possui um maior quantitativo de alunos, que estão sem esse dado preenchido. Esse reflexo é decorrente da não obrigatoriedade desse dado, já que somente a partir de 2017 foi homologado o parecer de número CNE/CEB Nº: 4/2017, que instituiu as Instituições de Ensino Básica e Superior a inclusão da raça/cor para o uso nos Censos Educacionais (MEC, 2017). A partir de então, a Unitins incluiu este atributo como obrigatório nos processos de matrícula para os alunos novatos e de rematrícula para os veteranos. Os demais atributos mesmo sendo importantes, observa-se que a maioria encontra-se na classe graduada.

## OS RECURSOS DO APRENDIZADO DE MÁQUINA

De acordo com MITCHELL (1997) o Aprendizado de Máquina (AM) é a construção de um sistema, capaz de adquirir conhecimentos de forma automática. Para os autores (MONARD e BARANAUSKAS, 2003), a hierarquia utilizada na maioria das vezes para a avaliação de um algoritmo de aprendizado é da Indução. A indução, pode ser entendida com a obtenção de conclusões genéricas a partir de um conjunto particular de exemplos. Para esta pesquisa utilizou-se o algoritmo de Floresta Aleatória (*Random Forests*) como método de AM na etapa de mineração de dados. Segundo BREIMAN (2001) estas árvores proporcionam uma precisão na classificação, por meio do crescimento de árvores que votam na classe mais popular.

Segundo MONTAÑO (2016), o método de Florestas Aleatórias é mais adequado para bases de dados que possuem diversas variações e requer poucos ajustes de parâmetros, caso da base de dados utilizada neste trabalho. Para o treinamento do algoritmo, a base de dados foi particionada em dois conjuntos aleatórios, um com 70% dos dados a serem usados no

treinamento, proporcional de acordo com as classes, e os demais 30% foram reservados para validação do treinamento.

Em se tratando da hierarquia do aprendizado de máquina, de acordo com os autores (MONARD e BARANAUSKAS, 2003), essa hierarquia do aprendizado geralmente é utilizada para avaliação de um algoritmo de aprendizado é da Indução. A indução, nada mais é a obtenção de conclusões genéricas a partir de um conjunto particular de exemplos. Ela é determinada como o raciocínio que se origina de um conceito específico e o generaliza, isto é, da parte para o todo. Assim sendo, a hipótese fornecida por meio da inferência indutiva podem ser verdadeiras ou não. Mesmo nesta condição, a inferência indutiva é um dos principais métodos conhecidos para advir o conhecimento novo ou prever eventos futuros. Sobretudo, o sistema de aprendizado indutivo ele é dividido em supervisionado e não-supervisionado. A escolha supervisionado ou não-supervisionado dependerá das tarefas e exemplos estarem ou não rotulados com o atributo de classe, especificamente essa pesquisa utilizou-se o aprendizado supervisionado.

A validação cruzada (*cross validation*), que é uma técnica para avaliar a capacidade de generalização de um modelo a partir de um conjunto de dados, foi utilizada nas etapas de treinamento e avaliação. Na validação cruzada é feito particionamento do conjunto de dados em subconjuntos mutuamente exclusivos, um elemento do conjunto é separado para testes (não participa do treinamento) e os demais elementos são utilizados no treinamento. Esse processo foi repetido  $k$  vezes alternando os elementos que foram utilizados para o grupo treinamento e para teste. A métrica de avaliação utilizada foi resultante das médias aritméticas dos  $k$  testes realizados (SCHAFFER, 1993).

Os algoritmos de aprendizado de máquina, em geral, possuem parâmetros que podem ser ajustados para um melhor desempenho, segundo alguma métrica. Na etapa de validação cruzada, foi utilizado o algoritmo chamado *Grid Search* que cria uma grade multidimensional com a combinação dos parâmetros a serem avaliados. Para o algoritmo de Florestas Aleatórias, dois parâmetros foram testados: o critério de entropia e a quantidade de estimadores (árvores da floresta). Os critérios de entropia foram o padrão (*entropy*) e o critério de Gini. Enquanto que os valores de estimadores foram testados num intervalo entre 10 (dez) e 120 (cento e vinte). Estes valores do intervalo foram obtidos de forma empírica e

em geral são dependentes do problema avaliado. Durante a etapa de validação cruzada, buscou-se obter a combinação de valores que maximizou o valor da acurácia, métrica que será explicada a seguir.

## **AS MÉTRICAS DA AVALIAÇÃO DA SOLUÇÃO PROPOSTA**

Em geral, sistemas de aprendizado de máquina são avaliados por meio de métricas pontuais, aplicadas tanto na base de dados de treinamento quanto de teste. De forma a subsidiar a discussão das métricas a serem adotadas para a solução proposta, a seguir, são explicadas duas métricas de modelos de predição considerando um problema com apenas duas classes (identificadas neste trabalho como positivo e negativo), que poderão ser utilizadas para avaliar e analisar os resultados gerados através da etapa de Mineração de Dados. Essas métricas são: Matriz de Confusão, que permite identificar a frequência das classes no modelo avaliado; A *Receiver Operating Characteristic* (ROC) e Área sob a Curva ROC ou *Area Under the Curve* (ROC-AUC), que serve para avaliar a qualidade do modelo.

A primeira maneira de se observar o desempenho de um classificador binário e através de uma Matriz de Confusão. Esta consiste de uma matriz quadrada onde são dispostas as previsões (linhas) e os valores verdadeiros (colunas). Ela é utilizada para mostrar a quantidade de acertos e erros de maneira que seja possível verificar a quantidade de amostras confundidas pelo sistema. Para tanto, é preciso definir o significado dos seguintes termos: Verdadeiro Positivo (VP), Falso Positivo (FP), Verdadeiro Negativo (VN) e Falso Negativo (FN) (RASCHKA, 2015).

Para o problema proposto neste trabalho, Verdadeiro Positivo (VP) é a quantidade de alunos que evadiram e o modelo identificou como evadido. O Verdadeiro Negativo (VN) é a quantidade de alunos que não evadiram e o modelo reportou como não evadido. Falso Positivo (FP) é a quantidade de alunos que não evadiram, mas foram erroneamente classificados como tal. Falso Negativo (FN) é a quantidade de alunos que evadiram, mas foram classificados como não evadido. Neste sentido, minimizar o FN é importante uma vez que, ao não identificar o aluno que evadiu, ações que poderiam evitar a evasão nestes casos podem deixar de ser tomadas.

Com os valores definidos de VP, FP, VN e FN é construída a matriz de confusão, a qual é exibida na Figura 3.

**Figura 3**

**Matriz de confusão com as duas classes (Evadido e Não Evadido)**

Condição Atual \ Predição	Evadido (P)	Não Evadido (N)
	Aluno Evadido (P)	VP (Verdadeiro Positivo)
Aluno não Evadido (N)	FN (evadido, mas não é extraído)	VN (Verdadeiro Negativo)

Fonte: Elaborado pelos autores com base em (RASCHKA, 2015)

A métrica acurácia ou exatidão (AC) diz o quanto o modelo acertou dentro de seu desempenho geral. A acurácia é obtida tomando-se a soma das previsões corretas, e dividindo-a pelo número total de previsões, respectivamente, conforme mostra a equação (1):

$$AC = \frac{VP+VN}{FP+FN+VP+VN} \quad (1)$$

Devido à bases de dados desbalanceadas, onde existem diferenças muito grandes entre o número de amostras de cada classe, em geral também utiliza-se outras métricas tais como Precisão e Revogação, F1-score curva ROC. Para dar suporte a estas métricas, é necessário estabelecer a Taxa de Falsos Positivos (FPR) e Taxa de Verdadeiro Positivo (TPR), que são definidas pelas equações (2) e (3), respectivamente. Estas taxas fornecem informações importantes, para as amostras positivas que foram identificadas corretamente no conjunto de total das amostras positivas.

$$FPR = \frac{FP}{N} = \frac{FP}{FP+VN} \quad (2)$$

$$TRP = \frac{VP}{P} = \frac{VP}{FN+VP} \quad (3)$$

Posto isto, define-se a Precisão (PRE) como a razão entre o número de verdadeiros positivos (VP) e a soma de Verdadeiro Positivos (VP) e Falso Positivos (FP), definida na equação (4).

A métrica PRE serve para medir se daqueles que foram classificados como positivos, efetivamente estão corretos. Já a revogação ou *recall* (REC), apresentada na equação (5), é quão completo os resultados estão, onde o REC de fato é o sinônimo de TPR. O REC pode ser realizado em situações que os FN são considerados mais prejudiciais do que os FP.

$$PRE = \frac{VP}{VP+FP} \quad (4)$$

$$REC = TPR = \frac{VP}{P} = \frac{VP}{FN+VP} \quad (5)$$

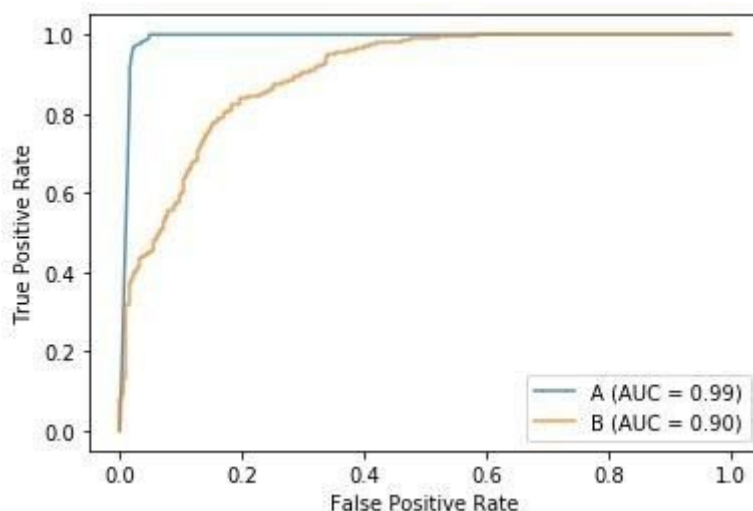
Para combinar as duas informações PRE e REC e uma medida só, há uma métrica chamada F1-Score, definida na equação (7). A F1-Score é considerada uma média harmônica entre duas medidas PRE e REC, permitindo que seja possível a identificação da qualidade geral do modelo, através de um único valor. Portanto, valores altos de F1 implicam em uma qualidade maior do modelo avaliado.

$$F1 = 2 \frac{PRE \times REC}{PRE+REC} \quad (6)$$

De acordo com (MARTÍNEZ-CAMBLOR e PARDO-FERNÁNDEZ, 2019), a curva de ROC é uma ferramenta utilizada na estatística para estudar a qualidade de um processo de classificação em termos de sensibilidade e especificidade. Quando a curva padrão estiver com os valores maiores no marcador de diagnóstico, significa que existe maior probabilidade de se manter as características estudadas. Um gráfico ROC é a relação entre TPR e FPR. A figura 4 apresenta a curva ROC gerada para dois classificadores A e B. No gráfico da figura 4, o classificador A é interpretado como sendo melhor do que B, segundo a curva ROC. Esta medida pode ser simplificada através do cálculo da área sob a curva, chamado de ROC-AUC, que naturalmente varia de 0 (zero) a 1 (um) (RASCHKA, 2015).

#### **Figura 4**

#### **Curva AUC (A e B)**



Fonte: (SOARES e colab., 2020)

Portanto, a avaliação do modelo de classificação pode ser feita por meio de diferentes métricas de acordo com as características dos dados. É importante que se leve em consideração, durante a avaliação de modelos de classificação, fatores como proporção de dados de cada classe e o objetivo da precisão.

## RESULTADOS

Tendo como base os treinamentos dos classificadores para os métodos aprendidos, apresenta-se os resultados obtidos por meio do Sistema Inteligente. As informações geradas pela solução inteligente, permitiu a identificação da predisposição da evasão escolar relacionada aos acadêmicos da Unitins. As métricas de avaliação empregadas para o modelo adotado foram: Acurácia, Precisão, Revogação, *F1-Score* e ROC-AUC.

Os resultados apresentados nas tabelas 4 e 5, foram gerados a partir de um modelo aprendido dos alunos evadidos e não evadidos, permitindo assim, a inferência da probabilidade do aluno evadir deste o primeiro período do curso. Para os resultados o modelo considerou apenas os alunos matriculados no segundo semestre de 2019, ou seja, um total de 1979 alunos ativos que estavam cursando disciplinas no semestre letivo de 2019-2.

**Tabela 4**

Propensão de evasão escolar por câmpus



<b>Câmpus</b>	<b>Qtde de Alunos Propensos a Evasão</b>	<b>Qtde de Alunos Predispostos a Formar</b>	<b>Total Alunos Matriculados</b>	<b>Propensão de Evasão</b>
Palmas	682	156	838	81,38%
Araguatins	147	94	241	60,69%
Augustinópolis	314	220	534	58,80%
Dianópolis	168	198	366	45,90%

Fonte: Elaborado pelos autores (2020)

**Tabela 5**

Ranqueamento dos 5 cursos com maiores índices de propensão de evasão

<b>Câmpus</b>	<b>Curso</b>	<b>Qtde de Alunos Propensos a Evasão</b>	<b>Qtde de Alunos Predispostos a Formar</b>	<b>Total Alunos Matriculados</b>	<b>Propensão de Evasão</b>
Palmas	Sistemas de Informação	127	10	137	92,7%
	Serviço Social	135	29	164	82,31%
	Direito	230	62	292	78,76%
	Engenharia Agrônômica	190	55	245	77,55%
Augustinópolis	Enfermagem	113	54	167	67,66%

Fonte: Elaborado pelos autores (2020)

Nota-se que o câmpus de Palmas é o maior em índices de chances de evasão, em seguida o câmpus de Araguaatins. Além disso, percebe-se que dos 12 cursos presenciais estudados, os cursos de Palmas são os que estão no *ranking* de evasão. Chama-se a atenção para o curso de Sistemas de Informação com a baixa probabilidade de graduação dos alunos matriculados neste curso.

Como forma de avaliar os resultados obtidos pelo Sistema Inteligente, apresenta-se a tabela 6 que expõe os resultados obtidos a partir do conjunto de treinamento que utilizou 70% da base

de dados. Para tal, foi utilizado o algoritmo de Floresta Aleatória pela técnica de validação. Observa-se que a média da acurácia entre os períodos é de 95,21% para os subconjuntos gerados pelo processo de validação cruzada, com subdivisão balanceada de 10 conjuntos. Isto significa que a parte de 70% foi dividida em 10 subconjuntos e portanto 10 rodadas foram feitas. Em cada rodada, 9 partes foram usadas para treinamento e 1 para testes, gerando-se portanto 10 classificadores de floresta aleatória. A Tabela 6 exibe as médias das métricas utilizadas nos 10 classificadores.

**Tabela 6**

**Resultados dos treinamentos a partir do 1º período dos cursos presenciais**

Métricas de Aprendizagem de Máquina	Resultados por período							
	1º Período	2º Período	3º Período	4º Período	5º Período	6º Período	7º Período	8º Período
Acurácia	93,9	94,3	94,1	94,7	94,8	96,2	96,5	97,2
Precisão	93,1	92,7	92,3	93,4	93,4	94,8	94,8	95,7
Revogação	96,5	97,8	97,8	97,6	97,8	98,7	99,3	99,6
F1	94,7	95,1	94,9	95,4	95,5	96,7	97	97,6
ROC-AUC	93,5	93,7	93,5	94,2	94,4	95,8	96,2	96,9

Fonte: Elaborado pelos autores (2020)

Para avaliação dos resultados foi utilizada a matriz de confusão, que está apresentada na tabela 7. A tabela 7 foi gerada por meio da validação cruzada e apresenta o quanto a metodologia utilizada nesta pesquisa foi eficaz em classificar corretamente os alunos evadidos e não evadidos. Os resultados evidenciam-se que a média dos períodos entre os Verdadeiro Negativo (VN) foi corretamente classificado em 78,1%. Isto significa que o classificador identificou corretamente os alunos não evadidos com uma precisão média de 78,1%. Já para os Verdadeiro Positivo (VP), chamado de aluno evadido a média de acertos foi de 53,1%. É

importante ressaltar que esta etapa serviu apenas para realizar o ajuste otimizado dos parâmetros (via um mecanismo chamado *Grid Search*) a serem utilizados no classificador final. O processo de *Grid Search* consiste em estipular um conjunto de parâmetros onde a combinação destes forma uma grid multidimensional. Cada ponto é uma combinação possível dos valores a serem explorados para otimização.

**Tabela 7**

**Matriz de Confusão gerada a partir do treinamento dos períodos dos cursos**

Matriz de Confusão - Validação Cruzada por Período	Resultados por período			
	VP	FP	FN	VN
<b>1º Período</b>	52,6	5,6	2,8	76,8
<b>2º Período</b>	52,1	6,1	1,7	77,9
<b>3º Período</b>	51,8	6,4	1,7	77,9
<b>4º Período</b>	52,8	5,4	1,9	77,7
<b>5º Período</b>	52,8	5,4	1,7	77,9
<b>6º Período</b>	54	4,2	1	78,6
<b>7º Período</b>	54	4,2	0,5	79,1
<b>8º Período</b>	54,7	3,5	0,3	79,3

Fonte: Elaborado pelos autores (2020)

A tabela 8 apresenta os resultados dos classificadores treinados por cada período, utilizando-se 70% de toda a base respectiva ao seu período. Os demais 30% foram utilizados

para testar os classificadores finais, respeitando-se a informação do período. O algoritmo utilizado foi de Floresta Aleatória, com seus respectivos parâmetros otimizados pelo processo de *Grid Search*. O treinamento foi realizado utilizando-se os 70% da base de dados. Para a extração das medidas de qualidade, para cada período foi testado com 30% da base. É importante ressaltar que os 30% de teste são desconhecidos do processo de treinamento e portanto são dados não vistos pelo classificador.

**Tabela 8**

**Resultado dos classificadores dos grupos de 70% de treinamento e 30% para teste**

Métricas de Aprendizagem de Máquina	Resultados por período							
	1º Período	2º Período	3º Período	4º Período	5º Período	6º Período	7º Período	8º Período
Acurácia	92,7	92,7	93,5	93,4	94,5	94,7	96,1	95,7
Precisão	91,4	90,2	92,2	91,7	92,5	92,3	93,9	93,9
Revogação	96,4	97,9	97	97,3	98,5	99,1	99,7	99,1
F1	93,8	93,9	94,5	94,4	95,4	95,6	96,7	96,4
ROC-AUC	92	91,7	92,9	92,6	93,8	93,8	95,4	95,1

Fonte: Elaborado pelos autores (2020)

O resultado gerado pela média entre os períodos para a acurácia é de 94,16%. Nota-se que os resultados obtidos na validação de 30% da base de dados, obtêm-se os resultados semelhantes com os alcançados durante o treinamento dos 70% dos dados. Pelos resultados, podemos concluir que o preditor se ajustou de tal forma, que conseguiu extrair o padrão representado por alunos evadidos e não-evadidos com uma precisão bastante elevada.

Outrossim, a tabela 9 apresenta a matriz de confusão com os resultados do grupo de 70% de treinamento e de 30% para testes. Essa matriz expressa a quantidade de acertos em classificar

a situação do acadêmico evadido ou não evadido. Nessa continuidade, verifica-se o bom desempenho dos métodos utilizados nesta pesquisa na obtenção de VP (alunos evadidos) e de VN (alunos não evadidos).

**Tabela 9**  
**Matriz de Confusão do grupo 70% de treinamento de 30% para teste**

Matriz de Confusão - Validação Cruzada por Período	Resultados por período Grupos de 70 e 30%			
	VP	FP	FN	VN
<b>1º Período</b>	218	31	12	330
<b>2º Período</b>	213	36	7	335
<b>3º Período</b>	221	28	10	332
<b>4º Período</b>	219	30	9	333
<b>5º Período</b>	222	27	5	337
<b>6º Período</b>	221	28	3	339
<b>7º Período</b>	227	22	1	341
<b>8º Período</b>	227	22	3	339

Fonte: Elaborado pelos autores (2020)

Observando os resultados apresentados na Tabela 9, percebe-se que nos 30% aleatoriamente selecionados, e conseqüentemente não utilizados durante a fase de treinamento, a quantidade de Falsos Negativos (FN) é zero. Observando a matriz de confusão exibida na Tabela 7, que remete aos dados de treinamento, percebe-se que este valor é menor que 1%. Ou seja, nos 70% dos dados, a validação cruzada separou em k=10 grupos, e o algoritmo foi treinado por

10 (dez) vezes seguidas, sempre utilizando-se 9 (nove) subgrupos para treinamento e 1 (um) para testes. Ainda neste cenário, o valor médio dos FN foi menor também próxima de zero.

É interessante ressaltar que o Falso Negativo representa o caso em que o aluno evadiu, mas foi classificado como não evadido. Este caso é o mais sensível para a aplicação considerada. Perceba que, se o sistema classificar um aluno provável de evadir como se não evadissem, os gestores do ensino não poderiam também considerar a situação deste aluno para criar elementos para tentar alcançá-lo, antes que este decida por evadir. Observando os Falsos Positivos, errar neste aspecto significaria errar por excesso, pois eles também seriam considerados para elaboração de políticas para lidar com o problema. Ainda analisando a Tabela 5 (cinco), percebe-se que 17 (dezesete) amostras se encaixam nesta situação. Isto é apenas uma proporção de 0,02 (dois centésimos), quando comparado com o total de amostras considerada na base de dados de validação. O excesso é mínimo e, portanto, suportável pelo poder informativo do experimento.

## **CONSIDERAÇÕES FINAIS**

A evasão escolar é um problema que afeta todos os níveis de ensino. Os aspectos relacionados a esse fenômeno são diversos e complexos e nem sempre estão associados ao âmbito acadêmico. O uso de ferramentas de inteligência computacional tais como técnicas de Aprendizado de Máquina, tem contribuído bastante para estudos sobre Evasão Escolar ao extrair, de bases de dados, informações e conhecimentos que podem subsidiar o processo de tomada de decisão das IES.

Este trabalho validou, por meio de um estudo de caso, a hipótese de que as técnicas de Aprendizado de Máquina podem ser usadas de maneira satisfatória em Mineração de Dados Educacionais, o que foi verificado por experimentos utilizando a base de dados acadêmica da Universidade Estadual do Tocantins (Unitins). Os resultados alcançados neste trabalho indicam que os métodos de Mineração de dados educacionais podem contribuir para o desenvolvimento de um classificador de confiança elevada, ou seja, capaz de automatizar a tarefa de classificar alunos em estado de evasão. Uma vez feita esta separação, é possível conduzir um estudo mais focado no grupo identificado e tentar encontrar padrões que levaram à evasão.

Sabe-se que os setores acadêmicos que lidam com diversos processos e, em geral, um grupo pequeno de trabalho é alocado na tarefa de tentar entender as razões da evasão nas instituições e propor soluções. O experimento computacional reportado nesta pesquisa dá fortes indicativos de que é possível aproximar um modelo de evasão a partir dos registros acadêmicos. Este trabalho lança luz ao uso de tecnologias para orientar políticas públicas para reverter os prejuízos gerados pela evasão escolar no âmbito universitário. Assim, foi verificado que modelos de Inteligência Artificial podem compor uma base de informações precisas e orientadas, permitindo assim uma gestão inteligente, focada e eficaz.

## REFERÊNCIAS BIBLIOGRÁFICAS

ADACHI, Ana Amélia Chaves Teixeira (2017). Evasão de estudantes de cursos de graduação da USP – Ingressantes nos anos de 2002, 2003 e 2004. 2017. 294p. Tese. (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo-SP, 2017.

BAGGI, Cristiane Aparecida Dos Santos e DOS SANTOS BAGGI, Cristiane Aparecida e LOPES, Doraci Alves (2011). Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. Avaliação: Revista da Avaliação da Educação Superior (Campinas). [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s1414-40772011000200007>>. , 2011

BARLEM, Jamila Geri Tomaschewski e colab. (2012). Opção e evasão de um curso de graduação em enfermagem: percepção de estudantes evadidos. Revista Gaúcha de Enfermagem. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s1983-14472012000200019>>. , 2012

BRASIL. Ministério da Educação. Secretaria de Educação Superior (1996a). Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. Acesso em: 11/04/2020. Disponível em: [http://www.andifes.org.br/wp-content/files\\_flutter/Diplomacao\\_Retencao\\_Evasao\\_Graduacao\\_em\\_IES\\_Publicas-1996.pdf](http://www.andifes.org.br/wp-content/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf).

BRASIL, Lei nº 9.394, de 20 de dezembro de 1996 (1996b). Estabelece as diretrizes e bases da educação nacional. Brasília, DF, dez 1996.

BREIMAN.L. (2001). Random forests. Machine Learning, 45:5–32, 2001

CARVALHO, Alessandro Pires (2017). Fatores institucionais associados à evasão na educação superior. 2017. 90 f. Dissertação. (Mestrado em Administração)-Programa de Pós-Graduação em Administração, Universidade Federal de Goiás, Goiânia.

COLPANI, Rogério (2019). Mineração de Dados Educacionais: um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar. Informática na educação: teoria & prática. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.22456/1982-1654.87880>>. , 2019

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. (1996); From Data Mining to Knowledge Discovery in Databases. Disponível em: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>. Acesso em: 28 fev 2020.

GAIOSO, N. P. L (2005). O fenômeno da evasão escolar na educação superior no Brasil. 2005. 75 f. Dissertação (Mestrado em Educação) Programa de Pós-Graduação em Educação da Universidade Católica de Brasília, Brasília, DF, 2005.

GOMES, E. C.; SOARES, D. B.; DESIDÉRIO, S. N.; ROCHA, A. S. DA R. S. DA. (2019). EVASÃO NO CURSO DE LICENCIATURA EM FÍSICA DA UNIVERSIDADE FEDERAL DO TOCANTINS: diagnóstico e primeiros resultados de um projeto de intervenção. Revista Observatório, v. 5, n. 5, p. 482-508, 1 ago.

GUIMARÃES, Orlineya Maciel; MARTINS, Eliana Canteiro Bolorino; LIMA, Maria Jose de Oliveira (2019). A EVASÃO NO ENSINO SUPERIOR: A UNESP CÂMPUS DE FRANCA - PERÍODO DE 2013-2018. CAMINE: Caminhos da Educação = Camine: Ways of Education, Franca, v. 11, n. 2, p. 136-161, mar. 2020. ISSN 2175-4217. Disponível em: <<https://periodicos.franca.unesp.br/index.php/caminhos/article/view/3013>>. Acesso em: 15 abr 2020.

HORTA, Cecília Eugenia Rocha. Associação Brasileira de Mantenedoras de Ensino Superior (2012). Evasão no ensino superior brasileiro. Disponível em: <https://abmes.org.br/arquivos/publicacoes/Cadernos25.pdf>. Acesso em: 27 abr 2020.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019). Sem desistências, número de graduados poderia dobrar no Brasil. [http://inep.gov.br/artigo/-/asset\\_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206](http://inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206). Acesso em 08 out 2019.

KIRA, L. P. (1998) A evasão no ensino superior: o caso do curso de pedagogia da Universidade Estadual de Maringá (1992-1996). Dissertação (Mestrado em Educação), Universidade Metodista de Piracicaba, 106 p.

LOBO, M. B. de C. M. (2012). Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. ABMES Cadernos. Brasília, set./dez. 2012.

[MARTÍNEZ-CAMBLOR](#), Pablo e [PARDO-FERNÁNDEZ](#), Juan C. (2019). Parametric estimates for the receiver operating characteristic curve generalization for non-monotone relationships. Statistical methods in medical research, v. 28, n. 7, p. 2032–2048, Jul 2019.



MARTINHO, Valquiria Ribeiro de Carvalho (2014). Sistema Inteligente para a predição de grupo de evasão discente. Disponível em: <https://repositorio.unesp.br/bitstream/handle/11449/100340/000751146.pdf?sequence=1&isAllowed=y>. Acesso em: 26 mai 2020.

MEC. Ministério da Educação (2017). PARECER HOMOLOGADO Despacho do Ministro, publicado no D.O.U. de 12/1/2018, Seção 1, Pág. 12. Disponível em: [http://portal.mec.gov.br/index.php?option=com\\_docman&view=download&alias=67801-pceb004-17-pdf&category\\_slug=julho-2017-pdf&Itemid=30192](http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=67801-pceb004-17-pdf&category_slug=julho-2017-pdf&Itemid=30192). Acesso em: 08 jun 2020.

MITCHELL, Tom M. (1997). Machine learning. 1997. Burr Ridge, IL: McGraw Hill, v. 45, 1997.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto (2003). Conceitos sobre Aprendizado de Máquina. Disponível em: <http://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>. Acesso em: 9 dez 2019.

MONTAÑO. RAZER (2016). Aplicação de Técnicas de Aprendizado de Máquina na Mensuração Florestal. Disponível em: <https://www.acervodigital.ufpr.br/bitstream/handle/1884/45346/R%20-%20T%20-%20RAZER%20ANTHOM%20NIZER%20ROJAS%20MONTANO.pdf?sequence=1&isAllowed=y>. Acesso em: 16 mar 2020.

PEREIRA, R. S.; ZAVALA, A. A.; SANTOS, A. C. (2011). Evasão na Universidade Federal de Mato Grosso. Revista de Estudos Sociais, v. 13, n. 26, p. 74-86, 2011.

RASCHKA, Sebastian (2015). Python Machine Learning. [S.l.]: Packt Publishing Ltd, 2015.

RIGO, S. J. ; Cambuzzi, w. ; Barbosa, J. L. V. ; CAZELLA, Sílvio (2014) . Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. Revista Brasileira de Informática na Educação. Disponível em: <https://www.br-ie.org/pub/index.php/rbie/article/view/2423>. Acesso em: 12 mai 2020

SABBATINI, M. (2015). Concepções e estratégias da aprendizagem participativa na educação a distância (EAD): contribuição das práticas dialógicas e comunicacionais para a autonomia discente. Revista Observatório, v. 1, n. 3, p. 80-99, 26 dez.

SANTOS JUNIOR, J. DA S.; REAL, G. C. M (2019). Fator institucional para a evasão na educação superior. Revista Internacional de Educação Superior, v. 6, p. e020037, 27 dez.

SCHAFFER. C. (1993). Selecting a classification method by cross-validation. Disponível em: <https://link.springer.com/content/pdf/10.1007/BF00993106.pdf>. Acesso em: 17 Mar 2020.

SCHARGEL, F. P.; SMINK, J. Estratégias para auxiliar o problema de evasão escolar. Tradução de Luiz Frazão Filho. Rio de Janeiro: Dunya, 2002

SOARES, L. C. C. P.; RONZANI, R. A.; CARVALHO, R. L.; DA SILVA, A. T. R. (2020). APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA EM UM CONTEXTO ACADÊMICO COM FOCO NA IDENTIFICAÇÃO DOS ALUNOS EVADIDOS E NÃO EVADIDOS. Disponível em: <https://revista.unitins.br/index.php/humanidadeseinovacao/article/view/3293>. Acesso em: 08 jun 2020.

SOUZA, Solange Lima de (2008). Evasão no Ensino Superior: Um estudo utilizando a mineração de dados como ferramenta de gestão do conhecimento em um banco de dados referente à graduação de Engenharia. Disponível em: <http://livros01.livrosgratis.com.br/cp064905.pdf>. Acesso: 11 abr 2020.

SOUZA, Thays Santos (2017). Estudo sobre a evasão em cursos de graduação presenciais na Universidade Federal de Goiás – UFG. 2017. 214 f. Dissertação. (Mestrado Profissional em Gestão Organizacional)-Programa de Pós-Graduação em Gestão Organizacional, Universidade Federal de Goiás, Catalão.

TINTO, V. (1975). Dropout from Higher Education: A theoretical synthesis of recent research. Review of Educational Research Winter, v. 45, n. 1, p. 89-125, 1975.

UNITINS (2020). Universidade Estadual do Tocantins. Graduação. Disponível em: <https://www.unitins.br/nportal/graduacao>. Acesso em: 11 abr 2020.

## 7 CONSIDERAÇÕES FINAIS

Neste estudo buscou-se a compreensão sobre o cenário da Evasão Escolar, especificamente no Ensino Superior Brasileiro, levando-se em consideração os estudos desenvolvidos por pesquisadores, o mapeamento das soluções tecnológicas no Brasil e no mundo e, além disso, que os dados acadêmicos da Unitins, podem ser transformados em informações úteis, subsidiando os tomadores de decisão em busca da redução da evasão.

Conforme verificado neste trabalho, a evasão é um fenômeno presente na Unitins, justificando-se a necessidade de soluções que promovam uma gestão proativa da evasão. Propor soluções que busquem contornar esse problema é o desafio de muitas Universidades. Mesmo que as Instituições de Ensino saibam suas taxas de evasão, predizer se o aluno irá evadir é algo muito complexo e o uso de ferramentas computacionais inteligentes, pode ser efetivo este tipo de tarefa, como foi mostrado nesta pesquisa. Outrossim, é uma via de mão dupla: a máquina prediz a probabilidade do aluno a evadir e o humano investiga os motivos na busca de prevenção.

O estudo de caso dos dados acadêmicos da Unitins permitiu a modelagem dos padrões de evasão que foi viabilizado por meio do uso da mineração de dados educacionais. Com isso, foi possível a aplicação da proposta metodológica de um Sistema Inteligente, capaz de identificar precocemente os alunos eminentes a evasão, por meio de relatórios estatísticos, com apontamentos de percentuais de evasão e graduação.

Em uma análise preliminar da evasão da Unitins, constatou-se que o Câmpus de Palmas é o que possui maiores quantitativos de alunos com chances de evasão. Isso significa, que dos 838 alunos matriculados em 2019-2 a probabilidade de estes alunos evadirem é de 81,38%. O *ranking* dos 5 cursos com maiores índices, percebe-se que 4 deles são de Palmas e 1 de Augustinópolis. O curso de Sistemas de Informação é o mais alarmante, a chance dos alunos matriculados a se formarem é de apenas 7,29%. Acredita-se que de posse deste resultado, a Instituição possa tomar medidas mais assertivas para a gestão da evasão. Vale ressaltar, que a metodologia adotada neste estudo, permite a inferência de um conjunto de dados da instituição, como por câmpus, cursos e alunos. Uma maneira de melhorar a mineração de dados é o uso de outras fontes de dados, como sócios-econômicos, ambientes virtuais de aprendizagem, questionários específicos e dentre outros, como forma de levar a

Instituição o entendimento das causas da evasão na Unitins, o que ficou fora do escopo da presente pesquisa.

## **7.1 Trabalhos Futuros**

Com base no estudo abordado neste trabalho, estabelecem-se algumas recomendações para pesquisas ou trabalhos futuros. Destes, os principais são:

- a) Prospecção tecnológica sobre tecnologias voltadas para este fenômeno, já que os números encontrados foram poucos representativos;
- b) A criação de uma base de armazenagem dos dados com os padrões que os algoritmos de mineração exigem, possibilitando a geração dos arquivos no formato apropriado para a Mineração de Dados Educacionais;
- c) Utilização de outras fontes de dados, para a identificação das razões da Evasão, por meio de dados sócios-econômicos, ambientes virtuais de aprendizagem, questionários específicos e dentre outros;
- d) A aplicação das metodologias apresentadas na pesquisa, em outras universidades;
- e) A utilização da Mineração de Dados Educacionais e Aprendizado de Máquina, como proposta para uso em outras questões relacionadas ao ambiente acadêmico.

## **7.2 Produções Bibliográficas**

O trabalho desenvolvido na linha de pesquisa dessa dissertação gerou algumas produções, tais como:

### **7.2.1 Registros de Programas de Computador**

Registro de programas de computador junto ao Instituto Nacional de Propriedade Industrial, sendo:

- a) SiMEU-Web
- b) SiMeu-Algoritmo

### **7.2.2 Artigos**

**a) Artigo Publicado**

SOARES, L. C. C. P.; RONZANI, R. A.; SILVA, A.T.R.; CARVALHO, R. L. APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA EM UM CONTEXTO ACADÊMICO COM FOCO NA IDENTIFICAÇÃO DOS ALUNOS EVADIDOS E NÃO EVADIDOS. HUMANIDADES & INOVAÇÃO, V. 7, P. 223, 2020.

**b) Artigo Submetido**

Soares, L. C. C. P.; Ronzani, R. A.; Silva, A.T.R.; Carvalho, R. L. Sistema Inteligente para Previsão de Evasão Escolar em um Contexto Universitário. DADOS – Revista de Ciências Sociais

**c) Capítulo de livro (submetido)**

Soares, L. C. C. P.; Silva, A.T.R.; Arruda, W.; Junior, P. R. G. F.; Carvalho, R. L. Evasão escolar com base em patentes e registros de softwares. Editora Universitária - EDUFT.

## REFERÊNCIAS

ABBAD, Gardênia e CARVALHO, Renata Silveira e ZERBINI, Thaís. Evasão em curso via internet: explorando variáveis explicativas. RAE eletrônica. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s1676-56482006000200008>>. , 2006

ADACHI, Ana Amélia Chaves Teixeira. Evasão de estudantes de cursos de graduação da USP – Ingressantes nos anos de 2002, 2003 e 2004. 2017. 294p. Tese. (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo-SP, 2017.

BAGGI, Cristiane Aparecida Dos Santos e DOS SANTOS BAGGI, Cristiane Aparecida e LOPES, Doraci Alves. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. Avaliação: Revista da Avaliação da Educação Superior (Campinas). [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s1414-40772011000200007>>. , 2011

BARROS, Renata Pitta e colab. Predição do rendimento dos alunos em lógica de programação com base no desempenho das disciplinas do primeiro período do curso de ciências e tecnologia utilizando técnicas de mineração de dados. Brazilian Journal of Development. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.34117/bjdv6n1-186>>. , 2020

BAKER, Ryan e ISOTANI, Seiji e CARVALHO, Adriana. Mineração de Dados Educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.5753/rbie.2011.19.02.03>>. , 2011

BRASIL. Ministério da Educação. Secretaria de Educação Superior (1996b). Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras. Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. Acesso em: 11/04/2020. Disponível em: [http://www.andifes.org.br/wp-content/files\\_flutter/Diplomacao\\_Retencao\\_Evasao\\_Graduacao\\_em\\_IES\\_Publicas-1996.pdf](http://www.andifes.org.br/wp-content/files_flutter/Diplomacao_Retencao_Evasao_Graduacao_em_IES_Publicas-1996.pdf)

BRASIL, Lei nº 9.394, de 20 de dezembro de 1996 (1996c). Estabelece as diretrizes e bases da educação nacional. Brasília, DF, dez 1996.

BRASIL, Lei nº 10.287, de 20 de Setembro de 2001. Altera dispositivo da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases da educação Nacional. Brasília, DF, set 2001.

BRITO, D. B, PASCOAL, T. A, ARAÚJO, J. G. G. O, LEMOS, M. O, RÊGO, T. G.(2015). Identificação de estudantes do primeiro semestre com risco de evasão através de técnicas de Data Mining. Disponível em: Acesso em: 25 nov 2019.

CABENA, Peter. Discovering data mining: from concept to implementation. [S.l.]: Prentice Hall, 1998.

CARVALHO, Alessandro Pires. Fatores institucionais associados à evasão na educação superior. 2017. 90 f. Dissertação. (Mestrado em Administração)-Programa de Pós-Graduação em Administração, Universidade Federal de Goiás, Goiânia.

COLPANI, Rogério. Mineração de Dados Educacionais: um estudo da evasão no ensino médio com base nos indicadores do Censo Escolar. Informática na educação: teoria & prática. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.22456/1982-1654.87880>>. , 2019

COSTA, Evandro e AGUIAR, Janderson e MAGALHÃES, Jonathas. Sistemas de Recomendação de Recursos Educacionais: conceitos, técnicas e aplicações. . [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.5753/cbie.jaie.2013.57>>. , 2013

CRESWELL, J. W. (2010). Projeto de pesquisa: Métodos Qualitativo, Quantitativo e Misto. 3ª edição. Porto Alegre. Editora Artmed.

DE MEDEIROS, Luciano Frontino. Inteligência Artificial Aplicada: UMA ABORDAGEM INTRODUTÓRIA. [S.l: s.n.], 2018.

DE SOUZA BRITO, Maria Tatiane e colab. Contribuições de um Plugin do tipo Report para a Identificação do Risco de Evasão no AVA Moodle com base em Visualização de Dados. Revista brasileira de historia / orgao da Associacao Nacional dos Professores Universitarios de Historia, ANPUH, v. 28, n. 0, p. 01, 16 Fev 2020. Acesso em: 22 abr 2020.

DUAN, Lian e DA XU, Li. Business Intelligence for Enterprise Systems: A Survey. IEEE Transactions on Industrial Informatics. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1109/tii.2012.2188804>>. , 2012.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. (1996); From Data Mining to Knowledge Discovery in Databases. Disponível em: <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1230>. Acesso em: 28 fev 2020.

FILHO, Roberto Leal Lobo e. Silva e colab. A evasão no ensino superior brasileiro. Cadernos de Pesquisa. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1590/s0100-15742007000300007>>. , 2007

GAIOSO, N. P. L. O fenômeno da evasão escolar na educação superior no Brasil. 2005. 75 f. Dissertação (Mestrado em Educação) Programa de Pós-Graduação em Educação da Universidade Católica de Brasília, Brasília, DF, 2005.

GANASCIA, Jean-Gabriel. Epistemology of AI Revisited in the Light of the Philosophy of Information. Knowledge, Technology & Policy. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.1007/s12130-010-9101-0>>. , 2010

GIL, Antônio Carlos. Métodos e técnicas de pesquisa social. [S.l: s.n.], 2008.

GOLDSCHMIDT, Ronaldo e PASSOS, Emmanuel. Data mining: um guia Prático. [S.l.]: Gulf Professional Publishing, 2005.

GOMES, E. C.; SOARES, D. B.; DESIDÉRIO, S. N.; ROCHA, A. S. DA R. S. DA. (2019). EVASÃO NO CURSO DE LICENCIATURA EM FÍSICA DA UNIVERSIDADE FEDERAL DO TOCANTINS: diagnóstico e primeiros resultados de um projeto de intervenção. Revista Observatório, v. 5, n. 5, p. 482-508, 1 ago.

GUIMARÃES, Orlineya Maciel; MARTINS, Eliana Canteiro Bolorino; LIMA, Maria Jose de Oliveira (2020). A EVASÃO NO ENSINO SUPERIOR: A UNESP CÂMPUS DE FRANCA - PERÍODO DE 2013-2018. CAMINE: Caminhos da Educação = Camine: Ways of Education, Franca, v. 11, n. 2, p. 136-161, mar. 2020. ISSN 2175-4217. Disponível em: <<https://periodicos.franca.unesp.br/index.php/caminhos/article/view/3013>>. Acesso em: 15 abr 2020.

HAN, Jiawei e PEI, Jian e KAMBER, Micheline. Data Mining: Concepts and Techniques. [S.l.]: Elsevier, 2011.

HIRJI, Karim K. Exploring data mining implementation. Communications of the ACM. [S.l.: s.n.]. Disponível em: <<http://dx.doi.org/10.1145/379300.379323>>. , 2001

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019a). Censo da Educação Superior 2018. Divulgação dos Resultados. Brasília- DF- 19 de Setembro de 2019.

INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2019b). Sem desistências, número de graduados poderia dobrar no Brasil. [http://inep.gov.br/artigo/-/asset\\_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206](http://inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/sem-desistencias-numero-de-graduados-poderia-dobrar-no-brasil/21206). Acesso em 08 out 2019.

JADRIC, M.; GARAČA, Ž.; ČUKUŠIĆ, M. Student dropout analysis with application of data mining methods. Journal of Contemporary Management, Toronto, v. 15, n. 1, p. 31-46, 2010.

KAMPFF, A. J. C, Ferreira, V. H, REATEGUI, E, LIMA, J. V. Identificação de Perfis de Evasão e Mau Desempenho para Geração de Alertas num Contexto de Educação a Distância. Disponível em: Acesso em: 15 nov 2019.

KIRA, L. P. A evasão no ensino superior: o caso do curso de pedagogia da Universidade Estadual de Maringá (1992-1996). Dissertação (Mestrado em Educação), Universidade Metodista de Piracicaba, 106 p, 1998.

KIPNIS, Bernardo. A pesquisa institucional e a educação superior brasileira: um estudo de caso longitudinal da evasão. Linhas Críticas, Brasília, v. 6, n. 1, p. 109-130, jul./dez. 2000.



KOHAVI, R. JOHN, G.H. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324.

KOHN, K , Moraes, C. H. O impacto das novas tecnologias na sociedade: conceitos e características da Sociedade da Informação e da Sociedade Digital. Disponível em: <https://www.intercom.org.br/papers/nacionais/2007/resumos/R1533-1.pdf>. Acesso em: 26 nov 2019.

MACHADO, Nilson José. Qualidade da educação: cinco lembretes e uma lembrança. *Estudos Avançados: Revista da USP, São Paulo*, n. 61, v. 21, p. 277-294, 2007.

MARTÍNEZ-CAMBLOR, Pablo e PARDO-FERNÁNDEZ, Juan C. Parametric estimates for the receiver operating characteristic curve generalization for non-monotone relationships. *Statistical methods in medical research*, v. 28, n. 7, p. 2032–2048, Jul 2019.

MARTINS, Claudia Aparecida. “Uma abordagem para pré-processamento de dados textuais em algoritmos de aprendizado”. [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.11606/t.55.2003.tde-08032004-164855>>. , [S.d.]

MCCARTHY, John e colab. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, v. 27, n. 4, p. 12–12, 15 Dez 2006. Acesso em: 8 nov 2019

MELO, Francisco. Area under the ROC Curve. *Encyclopedia of Systems Biology*. [S.l.]: Springer, New York, NY, 2013. p. 38–39. . Acesso em: 31 out 2019.

MITCHELL, Tom M. *Machine learning*. 1997. Burr Ridge, IL: McGraw Hill, v. 45, 1997

MONARD, M. C, BARANAUSKAS, J. A. Aplicações de Inteligência Artificial: Uma Visão Geral. Disponível em: <http://dcm.ffclrp.usp.br/~augusto/publications/2000-laptec.pdf>. Acesso em: 10 nov 2019.

MONARD, M. C, BARANAUSKAS, J. A.(2003). Conceitos sobre Aprendizado de Máquina. In: REZENDE, S.O. *Sistemas inteligentes: fundamentos e aplicações*. Barueri, SP: Manole, cap.4, p.89-94, 2003.

PEREIRA, R. S.; ZAVALA, A. A.; SANTOS, A. C. Evasão na Universidade Federal de Mato Grosso. *Revista de Estudos Sociais*, v. 13, n. 26, p. 74-86, 2011.

PITEIRA, Martinha e APARICIO, Manuela e COSTA, Carlos J. Ethics of Artificial Intelligence: Challenges. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI). [S.l: s.n.]. Disponível em: <<http://dx.doi.org/10.23919/cisti.2019.8760826>>. , 2019

RAMESH, V. e PARKAVI, P. e RAMAR, K. Predicting Student Performance: A Statistical and Data Mining Approach. International Journal of Computer Applications. [S.l.: s.n.]. Disponível em: <<http://dx.doi.org/10.5120/10489-5242>>. , 2013

RASCHKA, Sebastian. Python Machine Learning. [S.l.]: Packt Publishing Ltd, 2015

RUSSELL, Stuart J. e colab. Artificial Intelligence: A Modern Approach. [S.l.]: Prentice Hall, 2010.

RIGO, Sandro José e colab. Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. Revista Brasileira de Informática na Educação. [S.l.: s.n.]. Disponível em: <<http://dx.doi.org/10.5753/rbie.2014.22.01.132>>. , 2014

ROMERO, Cristóbal e colab. Introduction. Handbook of Educational Data Mining. [S.l.: s.n.]. Disponível em: <<http://dx.doi.org/10.1201/b10274-2>>. , 2010

SABBATINI, M. (2015). Concepções e estratégias da aprendizagem participativa na educação a distância (EAD): contribuição das práticas dialógicas e comunicacionais para a autonomia discente. Revista Observatório, v. 1, n. 3, p. 80-99, 26 dez.

SANTOS JUNIOR, J. DA S.; REAL, G. C. M (2019). Fator institucional para a evasão na educação superior. Revista Internacional de Educação Superior, v. 6, p. e020037, 27 dez.

SOUZA, Solange Lima de (2008). Evasão no Ensino Superior: Um estudo utilizando a mineração de dados como ferramenta de gestão do conhecimento em um banco de dados referente à graduação de Engenharia. Disponível em: <http://livros01.livrosgratis.com.br/cp064905.pdf>. Acesso: 11 abr 2020.

SOUZA, Thays Santos (2017). Estudo sobre a evasão em cursos de graduação presenciais na Universidade Federal de Goiás – UFG. 2017. 214 f. Dissertação. (Mestrado Profissional em Gestão Organizacional)-Programa de Pós-Graduação em Gestão Organizacional, Universidade Federal de Goiás, Catalão.

SOWA, J.. F. Conceptual Structures: Information Processing in Mind and Machine. New York, USA : Addison-Wesley Publishing Company, Inc., 1984.

SPADY, W. Dropouts from higher education: an interdisciplinary review and synthesis. Interchange, 1970, I, 64-85.

TINTO, V. Dropout from Higher Education: A theoretical synthesis of recent research. Review of Educational Research Winter, v. 45, n. 1, p. 89-125, 1975.

TURBAN, Efraim e colab. Business Intelligence: Um enfoque gerencial para a inteligência do negócio. [S.l.]: Bookman Editora, 2009.

UNITINS. Universidade Estadual do Tocantins. Lei nº 3.124, de 14 de julho de 2016. Transforma em autarquia a fundação que especifica, e adota outras providências. Disponível em:

<https://docs.google.com/viewerng/viewer?url=https://www.unitins.br/cms/Midia/Arquivos/LDAI9RTGFY3G6YGFY3F5TJZAIKW3MRBJLB39H5RICS.PDF>. Acesso em: 10 nov 2019.

UNITINS. Universidade Estadual do Tocantins (2019). Histórico de Implantação e Desenvolvimento da Unitins. Disponível em : <https://www.unitins.br/nportal/portal/page/show/historico>>. Acesso em: 21 out 2019.

UNITINS. Universidade Estadual do Tocantins (2020a). Graduação. Disponível em: <<https://www.unitins.br/nportal/graduacao>>. Acesso em: 22 abr 2020.

UNITINS. Universidade Estadual do Tocantins (2020b). Base de Dados. Acesso em: 05 mar 2020

UNITINS. Universidade Estadual do Tocantins (2020c). Regimento Acadêmico. Disponível em:

<https://docs.google.com/viewerng/viewer?url=https://www.unitins.br/cms/Midia/Arquivos/WIGA8VGEVPHNC5DRZTNBQTBUZG6F2QNW2TN9ERLVW0S.pdf>. Acesso em: 16 abr 2020.

## APÊNDICES

**Apêndice 1** - Comunicado de Registro do Programa de Computador do Sistema de Modelagem de Evasão Universitária (SiMEU - Web)

### FORMULÁRIO DE COMUNICADO DE PROGRAMA DE COMPUTADOR

INSTITUIÇÃO/INSTITUIÇÕES TITULAR(ES)	
Nome da Instituição	CNPJ
Universidade Federal do Tocantins	05.149.726/0001-04

Obs: Inserir linhas caso seja necessário adicionar mais instituições.

DADOS DO INVENTOR 1		
Nome completo: Leandra Cristina Cavina Piovesan Soares		Estado civil:
Data de nascimento:	Telefone:	
RG: Órgão emissor:	E-mail:	
CPF:	Profissão: I	
Endereço:		
Bairro:		Cidade:
CEP:		Estado:
Possui vínculo com a UFT: <input type="checkbox"/> Professor <input type="checkbox"/> Técnico administrativo <input type="checkbox"/> Aluno graduação <input type="checkbox"/> Aluno especialização <input checked="" type="checkbox"/> Aluno pós graduação		Possui vínculo com outra instituição: Não Nome da instituição:
RG:	Órgão Emissor:	Data de Emissão:

DADOS DO INVENTOR 2		
Nome completo: Robson Aparecido Ronzani		Estado civil:
Data de nascimento:	Telefone:	
RG:	E-mail:	
CPF:	Profissão:	
Endereço:		
Bairro:		Cidade:
CEP:		Estado:

<b>Possui vínculo com a UFT:</b> <input type="checkbox"/> Professor <input type="checkbox"/> Técnico administrativo <input type="checkbox"/> Aluno graduação <input checked="" type="checkbox"/> Aluno especialização <input type="checkbox"/> Aluno pós graduação		<b>Possui vínculo com outra instituição: Não</b> <b>Nome da instituição:</b>
<b>RG:</b>	<b>Órgão Emissor:</b>	<b>Data de Emissão:</b>

DADOS DO INVENTOR 3		
<b>Nome completo:</b> Rafael Lima de Carvalho		<b>Estado civil:</b>
<b>Data de nascimento:</b>	<b>Telefone:</b>	
<b>RG:</b>	<b>E-mail:</b>	
<b>CPF:</b>	<b>Profissão:</b>	
<b>Endereço:</b>		
<b>Bairro:</b>		<b>Cidade:</b>
<b>CEP:</b>		<b>Estado:</b>
<b>Possui vínculo com a UFT:</b> <input checked="" type="checkbox"/> Professor <input type="checkbox"/> Técnico administrativo <input type="checkbox"/> Aluno graduação <input type="checkbox"/> Aluno especialização <input type="checkbox"/> Aluno pós graduação		<b>Possui vínculo com outra instituição: Não</b> <b>Nome da instituição:</b>
<b>RG:</b>	<b>Órgão Emissor:</b>	<b>Data de Emissão:</b>

CARACTERIZAÇÃO DO PROGRAMA DE COMPUTADOR					
Título da dissertação, tese TCC ou projeto de pesquisa:	<b>SOLUÇÕES TECNOLÓGICAS PARA O PROBLEMA DA EVASÃO UNIVERSITÁRIA, SOB A ÓPTICA DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL</b>				
Título do Software	<b>SIMEU-WEB</b>				
Qual o problema técnico ela pretende resolver?	A evasão escolar e o longo tempo de estudos é um dos principais desafios para o sistema educacional Brasileiro. O sistema inteligente tem como principal objetivo, fornecer às Instituições de Ensino, por meio de consultas e relatórios a identificação dos alunos predispostos a formar e/ou evadir. Permitindo assim, uma ação proativa para minimizar o aluno em situação de abandono. O SIMEU-Web oferece uma interface web amigável de forma a receber os dados de entrada dos registros acadêmicos e produzir relatórios informando a porcentagem de evasão, de acordo com o modelo inteligente treinado através de florestas aleatórias.				
Linguagens:	<table border="1" style="width: 100%;"> <tr> <td><b>Plataforma:</b></td> <td><b>Web</b></td> </tr> <tr> <td><b>Framework:</b></td> <td><b>Django e Scikit-learn</b></td> </tr> </table>	<b>Plataforma:</b>	<b>Web</b>	<b>Framework:</b>	<b>Django e Scikit-learn</b>
<b>Plataforma:</b>	<b>Web</b>				
<b>Framework:</b>	<b>Django e Scikit-learn</b>				

	<b>Linguagem de programação:</b>	Python
	<b>Ferramentas:</b>	Spyder
	<b>Banco de Dados:</b>	Nenhum
Este programa é modificação tecnológica ou derivação? Caso afirmativo, informe o nome do programa original e respectivo número de registro:	<b>Não</b>	
Este programa é composto por outras obras de natureza intelectual?	<b>Não</b>	
Classificação do campo de aplicação:	<b>D01 -Ensin Regl Ensino regular</b> (pré-escolar, 1º grau, 2º grau, superior, pós-graduação, orientação profissional), <b>ED03-Adm/Pr Ens Instituição/Administração/Processo de ensino</b> (jardim escolar, escola maternal, jardim de infância, escola: de 1º grau, 2º grau, centro de ensino, de estudo supletivo, universidade, faculdade ou instituto superior de ensino, evasão escolar, serviços educacionais, equipamento escolar, método de ensino, didática: técnica de ensino, prática de ensino; ensino integrado, processo formal de ensino, processo não formal de ensino); <b>ED06-Educação</b> (pedagogia, ensino, sistema educacional, rede de ensino, educação de adulto, educação de base, de massa, etc, política educacional; educação ectra-escolar: educação comunitária, recuperadora).	
Classificação do tipo de programa:	<b>FA01-Ferrm Apoio Ferramenta de Apoio, IA01-Intlg Artf, IA02-Sist Especi Sistemas Especialistas</b>	
Informe a data de criação do programa:	<b>02/05/2020</b>	
<b>SIGILO E CONFIDENCIALIDADE</b>		
Já houve revelação para terceiros não vinculados ao NIT UFT?	<b>Não</b>	
Está em fase de publicação em algum periódico científico,	<b>Sim</b>	

congresso, tese, artigo ou resumo?	
Já foi revelada sob forma oral?	Não
<b>PERCENTUAL DE PARTICIPAÇÃO DOS AUTORES</b>	
Qual a quota parte ficou definida para cada autor, considerando o percentual de 30% em caso de transferência da tecnologia?	

Palmas, 04 de junho de 2020.

---

Assinatura de um dos autores

**Observações:**

1. Após preencher o presente formulário enviar para o e-mail [dituft@uft.edu.br](mailto:dituft@uft.edu.br) e entregar uma via assinada ao Núcleo de Inovação Tecnológica da UFT.
2. Núcleo de Inovação Tecnológica UFT, Avenida NS 15, ACNO 14, Bloco IV, Propesq, Sala 15, CEP 77001-090, Palmas/TO. Telefone: (63) 3229-4315 e-mail: dituft@uft.edu.br

## ANEXO I - VOCABULÁRIO

**Patenteabilidade** – verificação dos requisitos de novidade, atividade inventiva e aplicação industrial.

**Novidade** – aquilo que não está compreendido no estado da técnica.

**Estado da Técnica** - é constituído por tudo aquilo tornado acessível ao público antes da data de depósito do pedido de patente, por descrição escrita ou oral, por uso ou qualquer outro meio, no Brasil ou no exterior.

**Atividade Inventiva** – a invenção é dotada de atividade inventiva sempre que, para um técnico no assunto não decorra de maneira evidente ou óbvia do estado da técnica.

**Ato Inventivo** – o modelo de utilidade é dotado de ato inventivo sempre que, para um técnico no assunto, não decorra de maneira comum ou vulgar do estado da técnica.

**Aplicação Industrial** – quando pode ser utilizado ou produzido em qualquer tipo de indústria.

**Invenção** – é todo produto ou equipamento, ou mesmo processo novo, de cunho científico ou tecnológico, cuja criatividade seja considerada pioneira para o conhecimento público, que não seja evidente ou óbvia para a comunidade científica e que seja susceptível de aplicação ou produção industrial.

**Modelo de Utilidade** – objeto de uso prático, ou parte deste, susceptível de aplicação industrial que apresente nova forma ou disposição, envolvendo ato inventivo, que resulte em melhoria funcional no seu uso ou em sua fabricação.

**Termo de Sigilo** – Instrumento utilizado para garantir a confidencialidade das informações relacionadas à pesquisa.

**Desenho Industrial** – a forma plástica ornamental de um objeto ou o conjunto ornamental de linhas e cores que possa ser aplicado a um produto, proporcionando resultado visual novo e original na sua configuração externa e que possa servir de tipo de fabricação industrial.

**Programa de Computador** – é a expressão de um conjunto organizado de instruções em linguagem natural ou codificada contida em suporte físico de qualquer natureza, de emprego necessário em máquinas automáticas de tratamento da informação, dispositivos, instrumentos ou equipamentos periféricos, baseados em técnica digital ou análoga, para fazê-los funcionar de modo e para fins determinados.

**Direitos Autorais** – os direitos do autor e os que lhes são conexos.

**Obras Intelectuais Protegidas** – as criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro.

**Apêndice 2** - Comunicado de Programa de Computador do Sistema de Modelagem de Evasão Universitária (SiMEU)

## FORMULÁRIO DE COMUNICADO DE PROGRAMA DE COMPUTADOR

INSTITUIÇÃO/INSTITUIÇÕES TITULAR(ES)	
Nome da Instituição	CNPJ
Universidade Federal do Tocantins	05.149.726/0001-04

Obs: Inserir linhas caso seja necessário adicionar mais instituições.

DADOS DO INVENTOR 1	
Nome completo: Leandra Cristina Cavina Piovesan Soares	Estado civil:
Data de nascimento:	Telefone:
RG: Órgão emissor:	E-mail:
CPF:	Profissão:
Endereço:	



<b>Bairro:</b>		<b>Cidade:</b>
<b>CEP:</b>		<b>Estado:</b>
<b>Possui vínculo com a UFT:</b> <input type="checkbox"/> Professor <input type="checkbox"/> Técnico administrativo <input type="checkbox"/> Aluno graduação <input type="checkbox"/> Aluno especialização <input checked="" type="checkbox"/> Aluno pós graduação		<b>Possui vínculo com outra instituição: Não</b> <b>Nome da instituição:</b>
<b>RG:</b>	<b>Órgão Emissor:</b>	<b>Data de Emissão:</b>

DADOS DO INVENTOR 2		
<b>Nome completo: Robson Aparecido Ronzani</b>		<b>Estado civil:</b>
<b>Data de nascimento:</b>	<b>Telefone:</b>	
<b>RG:</b>	<b>E-mail:</b>	
<b>CPF:</b>	<b>Profissão:</b>	
<b>Endereço:</b>		
<b>Bairro:</b>		<b>Cidade:</b>
<b>CEP:</b>		<b>Estado:</b>
<b>Possui vínculo com a UFT:</b> <input type="checkbox"/> Professor <input type="checkbox"/> Técnico administrativo <input type="checkbox"/> Aluno graduação <input checked="" type="checkbox"/> Aluno especialização <input type="checkbox"/> Aluno pós graduação		<b>Possui vínculo com outra instituição: Não</b> <b>Nome da instituição:</b>
<b>RG:</b>	<b>Órgão Emissor:</b>	<b>Data de Emissão:</b>

DADOS DO INVENTOR 3		
<b>Nome completo: Rafael Lima de Carvalho</b>		<b>Estado civil:</b>
<b>Data de nascimento:</b>	<b>Telefone:</b>	
<b>RG:</b>	<b>E-mail:</b>	
<b>CPF:</b>	<b>Profissão:</b>	
<b>Endereço:</b>		
<b>Bairro:</b>		<b>Cidade:</b>
<b>CEP:</b>		<b>Estado:</b>
<b>Possui vínculo com a UFT:</b> <input checked="" type="checkbox"/> Professor <input type="checkbox"/> Técnico administrativo <input type="checkbox"/> Aluno graduação <input type="checkbox"/> Aluno especialização <input type="checkbox"/> Aluno pós graduação		<b>Possui vínculo com outra instituição: Não</b> <b>Nome da instituição:</b>

<b>RG:</b>	<b>Órgão Emissor:</b>	<b>Data de Emissão:</b>
------------	-----------------------	-------------------------

<b>CARACTERIZAÇÃO DO PROGRAMA DE COMPUTADOR</b>									
Título da dissertação, tese TCC ou projeto de pesquisa:	<b>SOLUÇÕES TECNOLÓGICAS PARA O PROBLEMA DA EVASÃO UNIVERSITÁRIA, SOB A ÓPTICA DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL</b>								
Título do Software	<b>Sistema de Modelagem de Evasão Universitária (SiMEU)</b>								
Qual o problema técnico ela pretende resolver?	A evasão escolar é um dos principais desafios atuais das Instituições de Ensino Superior. O SiMeu é uma solução inteligente, que utiliza as técnicas de Mineração de dados educacionais e aprendizado de máquina, treinado por meio de florestas aleatórias. Esta solução visa a identificação dos alunos evadidos e não evadidos que estão no banco de dados e a partir deste modelo construído a predição dos alunos com propensão a evasão.								
Linguagens:	<table border="1"> <tr> <td><b>Framework:</b></td> <td><b>Scikit-learn e pandas</b></td> </tr> <tr> <td><b>Linguagem de programação:</b></td> <td><b>Python</b></td> </tr> <tr> <td><b>Ferramentas:</b></td> <td><b>Spyder</b></td> </tr> <tr> <td><b>Banco de Dados:</b></td> <td><b>Nenhum</b></td> </tr> </table>	<b>Framework:</b>	<b>Scikit-learn e pandas</b>	<b>Linguagem de programação:</b>	<b>Python</b>	<b>Ferramentas:</b>	<b>Spyder</b>	<b>Banco de Dados:</b>	<b>Nenhum</b>
<b>Framework:</b>	<b>Scikit-learn e pandas</b>								
<b>Linguagem de programação:</b>	<b>Python</b>								
<b>Ferramentas:</b>	<b>Spyder</b>								
<b>Banco de Dados:</b>	<b>Nenhum</b>								
Este programa é modificação tecnológica ou derivação? Caso afirmativo, informe o nome do programa original e respectivo número de registro:	<b>Não</b>								
Este programa é composto por outras obras de natureza intelectual?	<b>Não</b>								
Classificação do campo de aplicação:	<b>D01 -Ensin Regl Ensino regular</b> (pré-escolar, 1º grau, 2º grau, superior, pós-graduação, orientação profissional), <b>ED03-Adm/Pr Ens Instituição/Administração/Processo de ensino</b> (jardim escolar, escola maternal, jardim de infância, escola: de 1º grau, 2º grau, centro de ensino, de estudo supletivo, universidade, faculdade ou instituto superior de ensino, evasão escolar, serviços educacionais, equipamento escolar, método de ensino, didática: técnica de ensino, prática de ensino; ensino integrado, processo formal de ensino, processo não formal de ensino); <b>ED06-Educação</b> (pedagogia, ensino, sistema educacional, rede de ensino, educação de adulto, educação de base, de massa, etc, política educacional; educação ectra-escolar: educação comunitária, recuperadora).								
Classificação do tipo de programa:	<b>FA01-Ferrm Apoio Ferramenta de Apoio, IA01-Intlg Artf, IA02-Sist Especl Sistemas Especialistas</b>								

Informe a data de criação do programa:	15/04/2020
<b>SIGILO E CONFIDENCIALIDADE</b>	
Já houve revelação para terceiros não vinculados ao NIT UFT?	Não
Está em fase de publicação em algum periódico científico, congresso, tese, artigo ou resumo?	Sim
Já foi revelada sob forma oral?	Não
<b>PERCENTUAL DE PARTICIPAÇÃO DOS AUTORES</b>	
Qual a quota parte ficou definida para cada autor, considerando o percentual de 30% em caso de transferência da tecnologia?	

Palmas, 17 de junho de 2020.

---

Assinatura de um dos autores

**Observações:**

1. Após preencher o presente formulário enviar para o e-mail [dituft@uft.edu.br](mailto:dituft@uft.edu.br) e entregar uma via assinada ao Núcleo de Inovação Tecnológica da UFT.
2. Núcleo de Inovação Tecnológica UFT, Avenida NS 15, ACNO 14, Bloco IV, Propesq, Sala 15, CEP 77001-090, Palmas/TO. Telefone: (63) 3229-4315 e-mail: [dituft@uft.edu.br](mailto:dituft@uft.edu.br)

## ANEXO I - VOCABULÁRIO

**Patenteabilidade** – verificação dos requisitos de novidade, atividade inventiva e aplicação industrial.

**Novidade** – aquilo que não está compreendido no estado da técnica.

**Estado da Técnica** - é constituído por tudo aquilo tornado acessível ao público antes da data de depósito do pedido de patente, por descrição escrita ou oral, por uso ou qualquer outro meio, no Brasil ou no exterior.

**Atividade Inventiva** – a invenção é dotada de atividade inventiva sempre que, para um técnico no assunto não decorra de maneira evidente ou óbvia do estado da técnica.

**Ato Inventivo** – o modelo de utilidade é dotado de ato inventivo sempre que, para um técnico no assunto, não decorra de maneira comum ou vulgar do estado da técnica.

**Aplicação Industrial** – quando pode ser utilizado ou produzido em qualquer tipo de indústria.

**Invenção** – é todo produto ou equipamento, ou mesmo processo novo, de cunho científico ou tecnológico, cuja criatividade seja considerada pioneira para o conhecimento público, que não seja evidente ou óbvia para a comunidade científica e que seja susceptível de aplicação ou produção industrial.

**Modelo de Utilidade** – objeto de uso prático, ou parte deste, susceptível de aplicação industrial que apresente nova forma ou disposição, envolvendo ato inventivo, que resulte em melhoria funcional no seu uso ou em sua fabricação.

**Termo de Sigilo** – Instrumento utilizado para garantir a confidencialidade das informações relacionadas à pesquisa.

**Desenho Industrial** – a forma plástica ornamental de um objeto ou o conjunto ornamental de linhas e cores que possa ser aplicado a um produto, proporcionando resultado visual novo e original na sua configuração externa e que possa servir de tipo de fabricação industrial.

**Programa de Computador** – é a expressão de um conjunto organizado de instruções em linguagem natural ou codificada contida em suporte físico de qualquer natureza, de emprego necessário em máquinas automáticas de tratamento da informação, dispositivos, instrumentos ou equipamentos periféricos, baseados em técnica digital ou análoga, para fazê-los funcionar de modo e para fins determinados.

**Direitos Autorais** – os direitos do autor e os que lhes são conexos.

**Obras Intelectuais Protegidas** – as criações do espírito, expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível, conhecido ou que se invente no futuro.

# ANEXOS

## Anexo 1 - Aprovação da Plataforma Brasil

23/07/2020

Plataforma Brasil

BRASIL



principal



sair

LEANDRA CRISTINA CAVINA PIOVESAN SOARES - | V3.2

Sua sessão expira em: 39min 27

Você está em: Público > Buscar Pesquisas Aprovadas > Detalhar Projeto de Pesquisa

DETALHAR PROJETO DE PESQUISA	
<b>DADOS DO PROJETO DE PESQUISA</b>	
Título Público: SOLUÇÕES TECNOLÓGICAS PARA O PROBLEMA DA EVASÃO UNIVERSITÁRIA, SOB A ÓPTICA DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL	
Pesquisador Responsável: LEANDRA CRISTINA CAVINA PIOVESAN SOARES	
Contato Público: LEANDRA CRISTINA CAVINA PIOVESAN SOARES	
Condições de saúde ou problemas estudados:	
Descritores CID - Gerais:	
Descritores CID - Específicos:	
Descritores CID - da Intervenção:	
Data de Aprovação Ética do CEP/CONEP: 10/03/2020	
<b>DADOS DA INSTITUIÇÃO PROPONENTE</b>	
Nome da Instituição: UNIVERSIDADE ESTADUAL DO TOCANTINS - UNITINS	
Cidade: PALMAS	
<b>DADOS DO COMITÊ DE ÉTICA EM PESQUISA</b>	
Comitê de Ética Responsável: 8023 - UNIVERSIDADE ESTADUAL DO TOCANTINS - UNITINS	
Endereço: 108 Sul, Alameda 11, Lote 3	
Telefone: (63)3218-2929	
E-mail: cep@unitins.br	
<b>CENTRO(S) PARTICIPANTE(S) DO PROJETO DE PESQUISA</b>	<input type="text"/>
<b>CENTRO(S) COPARTICIPANTE(S) DO PROJETO DE PESQUISA</b>	<input type="text"/>
<a href="#">Voltar</a>	

Suporte a sistemas: 136 - opção 8 e opção 3, solicitar ao atendente suporte Plataforma Brasil.  
Fale conosco: Clique para enviar mensagem para a Plataforma Brasil

