

UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE EDUCAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS  
EDUCACIONAIS EM REDE – MESTRADO PROFISSIONAL

Thiago Siqueira Sonnenstrahl

**UTILIZAÇÃO DA MINERAÇÃO DE DADOS PARA IDENTIFICAR A  
EVASÃO NOS CURSOS EAD DO INSTITUTO FEDERAL DE  
EDUCAÇÃO, CIÊNCIA E TECNOLOGIA FARROUPILHA**

Santa Maria, RS

2020

**Thiago Siqueira Sonnenstrahl**

**UTILIZAÇÃO DA MINERAÇÃO DE DADOS PARA IDENTIFICAR A EVASÃO  
NOS CURSOS EAD DO INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E  
TECNOLOGIA FARROUPILHA**

Dissertação apresentada ao Programa de Pós-Graduação em Tecnologias Educacionais em Rede, Área de Concentração em Tecnologias Educacionais em Rede para Inovação e Democratização da Educação, da Universidade Federal de Santa Maria, como requisito parcial para a obtenção do título de **Mestre em Tecnologias Educacionais em Rede**.

Orientadora: Prof. Dra. Solange de Lurdes Pertile

Santa Maria, RS

2020

Sonnenstrahl, Thiago Siqueira

Utilização da mineração de dados para identificar a evasão nos cursos EaD do Instituto Federal de Educação, Ciência e Tecnologia Farroupilha / Thiago Siqueira  
Sonnenstrahl.- 2020.

81 p.; 30 cm

Orientadora: Solange Pertile

Coorientadora: Giliane Bernardi

Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Educação, Programa de Pós-Graduação em Tecnologias Educacionais em Rede, RS, 2020

1. Mineração de Dados Educacionais 2. Evasão I.  
Pertile, Solange II. Bernardi, Giliane III. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

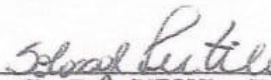
Declaro, THIAGO SIQUEIRA SONNENSTRAHL, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Dissertação) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

**Thiago Siqueira Sonnenstrahl**

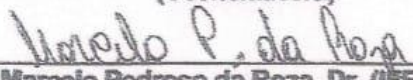
**UTILIZAÇÃO DA MINERAÇÃO DE DADOS PARA IDENTIFICAR A EVASÃO  
NOS CURSOS EAD DO INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E  
TECNOLOGIA FARROUPILHA**

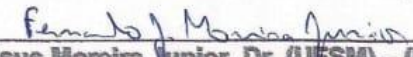
Dissertação apresentada ao Programa de Pós-Graduação em Tecnologias Educacionais em Rede, Área de Concentração em Tecnologias Educacionais em Rede para Inovação e Democratização da Educação, da Universidade Federal de Santa Maria, como requisito parcial para a obtenção do título de Mestre em Tecnologias Educacionais em Rede.

**Aprovada em 17 de março de 2020:**

  
\_\_\_\_\_  
**Solange de Lurdes Pertile, Dra. (UFSM) – (Por Videoconferência)**  
(Presidente/Orientadora)

  
\_\_\_\_\_  
**Gilliane Bernardi, Dra. (UFSM) – (Por Videoconferência)**  
(Coorientadora)

  
\_\_\_\_\_  
**Marcelo Pedroso da Roza, Dr. (IFFar AL)**

  
\_\_\_\_\_  
**Fernando de Jesus Moreira Junior, Dr. (UFSM) – (Por Videoconferência)**

Santa Maria, RS

2020

## AGRADECIMENTOS

Primeiramente, agradeço minha família que sempre apoiou minha qualificação e meu crescimento profissional, sendo minha base, dispondo de todo carinho e suporte em todos momentos da minha vida.

Ao Fernando Lucas Oliveira, colega e amigo do Instituto Federal de Sergipe, que contribuiu no projeto de mestrado e muito auxiliou no início desta caminhada.

À Miriam Pizzatto Colpo, colega de sala e amiga, por torcer pelo meu sucesso desde o início, pelas inúmeras sugestões neste trabalho, sanando meus principais anseios, compartilhando do seu conhecimento e experiência, sendo uma pessoa fundamental para a realização desta pesquisa, e, também, pela compreensão nos dias mais difíceis no trabalho.

Ao Cristiano, conhecido por “jiló”, pela torcida e pelo apoio desde a inscrição no mestrado, pelo apoio ao projeto e por sanar minhas dúvidas no decorrer do trabalho.

À Suelen, minha namorada, por estar ao meu lado incentivando nos momentos difíceis, para que eu pudesse retomar a tranquilidade e concluir o mestrado.

Aos colegas e amigos que apoiaram e foram solícitos sempre que precisei, Rafael e José; ao Diego, ao Juliano e à Cristiane, estes colegas de mestrado e de IF. À Naura e à Patrícia, também, colegas do mestrado, com quem tive a oportunidade de compartilhar, além dos trabalhos do mestrado, risadas, experiências e aprendizados.

À orientadora, Solange Pertile, pelos ensinamentos, pelas contribuições, pela paciência e pelas sugestões apontadas desde o início desta caminhada.

À coorientadora, Giliane Bernardi, pelo incentivo, pelas conversas, pela disponibilidade e pelas considerações realizadas nesta pesquisa.

À Diretoria de educação a distância, em especial à Monique, pelas considerações e análises nesta dissertação; ao Bruno, pela disposição em sanar qualquer dúvida com relação ao banco de dados e demais informações sempre que necessário; ao André Dias, por conseguir o acesso ao banco de dados.

À banca examinadora, Marcelo e Fernando, pelas ideias, contribuições e sugestões desde a qualificação desta dissertação.

## RESUMO

### UTILIZAÇÃO DA MINERAÇÃO DE DADOS PARA IDENTIFICAR A EVASÃO NOS CURSOS EAD DO INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA FARROUPILHA

AUTOR: Thiago Siqueira Sonnenstrahl  
ORIENTADORA: Solange Pertile

O Instituto Federal Farroupilha, como uma componente da Rede Federal de Educação Básica, Profissional, Técnica e Tecnológica, tem a permanência e o êxito dos estudantes como uma das metas do Plano de Desenvolvimento Institucional (PDI) 2019/2026. Gerenciar o desempenho de alunos em um ambiente virtual de ensino e aprendizagem (AVEA) é de fundamental importância para a redução dos índices de evasão e reprovação nos cursos da modalidade de Ensino a Distância (EaD). Assim, esta pesquisa tem como objetivo, através da Mineração de Dados Educacionais (MDE), analisar, por meio da interação dos alunos no AVEA, possíveis evasões em cursos do Instituto Federal Farroupilha na modalidade a distância, disponibilizando dados estratégicos para os gestores educacionais da instituição. O desenvolvimento do trabalho dividiu-se em quatro etapas distintas, baseando seu procedimento em uma pesquisa bibliográfica, juntamente a uma abordagem quali-quantitativa. A primeira etapa buscou, por meio de uma pesquisa exploratória, dados de evasão e demais informações junto à Diretoria de educação a distância do Instituto Federal Farroupilha (IFFar). A segunda etapa deu-se com uma revisão bibliográfica acerca do estudo da evasão no EaD. A terceira etapa foi a de mineração de dados e avaliação dos resultados. A quarta e última etapa consistiu-se de uma análise qualitativa dos dados da mineração, como forma de basear a instituição para tomada de decisão no âmbito da Diretoria de Educação a Distância, considerando-se a interação dos alunos no AVEA. O desenvolvimento da pesquisa foi realizado por meio de três experimentos, utilizando interações no AVEA Moodle de duas turmas de um curso subsequente na modalidade EaD. Cada experimento consistiu em uma turma, e o terceiro experimento foi a unificação dos dados em um único conjunto. Como resultado, na mineração do experimento 3, que uniu os dados das duas turmas, a taxa de acerto foi superior a 88%, obtido com o algoritmo *Randon Forest*. Os melhores atributos que realizaram a predição foram visualização de tarefa e visualização de material. A dissertação de mestrado apresentada está inserida na linha de pesquisa de Desenvolvimento de Tecnologia Educacional em Rede, do Programa de Pós-Graduação em Tecnologias Educacionais em Rede, e gerou como produtos o próprio texto aqui apresentado e a estratégia de MDE criada.

**Palavras-chave:** Educação a Distância. Evasão. Mineração de Dados Educacionais.

## ABSTRACT

### USE OF DATA MINING TO IDENTIFY DROPOUT RATES OF DE COURSES OF THE FARROUPILHA FEDERAL INSTITUTE OF EDUCATION, SCIENCE, AND TECHNOLOGY

AUTHOR: Thiago Siqueira Sonnenstrahl  
ADVISOR: Solange Pertile

The Farroupilha Federal Institute is a component of the Federal Network of Basic, Professional, Technical, and Technological Education and strives for the presence and success of its students in accordance with Institutional Development Plan (IDP) 2019/2026. Managing the performance of students in a virtual teaching and learning environment (VLE) is of fundamental importance to reduce dropout and failure rates in distance education (DE) courses. Thus, by using Educational Data Mining (EDM) and assessing student interaction on the VTLE, this study aimed to analyze possible dropouts in DE courses at the Farroupilha Federal Institute by providing strategic data for educational managers of the institution. The development of the present study was divided into four distinct stages and based on a bibliographic review employing a qualitative and quantitative approach. The first stage sought, through exploratory research, dropout data and other information from the distance education department of the Farroupilha Federal Institute. The second stage took place with a bibliographic review on dropout rates in distance education. The third step was data mining and the evaluation of results. The fourth and last stage consisted of a qualitative analysis of mining data as a way of guiding the institution to make decisions within the scope of the Distance Education Department while considering student interactions on the VTLE. The study was developed by performing three experiments using interactions on the VLE Moodle of two classes of a subsequent distance education course. Each experiment consisted of a class and the third experiment was the unification of the data in a single set. As a result, the mining of experiment 3, which joined the data of both classes and was obtained with the *Random Forest* algorithm, showed that the score rate was higher than 88%. The best attributes that performed the prediction were task visualization and material visualization. The master's dissertation presented here is in the line of research of the Development of Educational Technology in Networks, part of the Graduate Program in Educational Technology in Networks and generated as products the text presented here and the created EDM strategy.

**Keywords:** Distance Education. Dropout. Educational Data Mining.

## LISTA DE FIGURAS

Figura 1 – Interface Moodle no IFFar .....	30
Figura 2 – Etapas do processo de KDD .....	32
Figura 3 – Interface gráfica do WEKA .....	38
Figura 4 – Etapas de desenvolvimento da pesquisa .....	45
Figura 5 – Interface web do SISTEC .....	48
Figura 6 – Dados preparados.....	49
Figura 7 – Arquivo ARFF.....	50
Figura 8 – Árvore de decisão em porcentagem: primeiro experimento .....	52
Figura 9 – Árvore de decisão em porcentagem: segundo experimento .....	58
Figura 10 – Árvore de decisão em porcentagem: terceiro experimento .....	63
Figura 11 – Árvore de decisão com os dados discretizados .....	64



## LISTA DE QUADROS

Quadro 1 – Fatores internos e externos da evasão .....	23
Quadro 2 – Motivos da evasão EaD.....	25
Quadro 3 – Análise dos trabalhos correlatos.....	43
Quadro 4 – Dados brutos extraídos do AVEA .....	47
Quadro 5 – Atributos do AVEA Moodle .....	47
Quadro 6 – Atributos e dados finais utilizados na MDE .....	50

## LISTA DE TABELAS

Tabela 1 – Incidência das categorias de atividades que contribuíram para a permanência.....	24
Tabela 2 – Métricas de desempenho dos algoritmos utilizados: primeiro experimento .....	51
Tabela 3 – Interação máxima: primeiro experimento .....	52
Tabela 4 – Métricas de desempenho dos algoritmos utilizados: segundo experimento.....	56
Tabela 5 – Interação máxima: segundo experimento.....	57
Tabela 6 – Métricas de desempenho dos algoritmos utilizados: terceiro experimento .....	62
Tabela 7 – Interação máxima: terceiro experimento .....	62

## LISTA DE GRÁFICOS

Gráfico 1 – Número de matrículas na EaD ao longo do tempo .....	18
Gráfico 2 – Polos EaD .....	18
Gráfico 3 – Interações totais nos recursos utilizados do AVEA: primeiro experimento .....	54
Gráfico 4 – Média de utilização dos recursos no AVEA: Primeiro experimento ....	55
Gráfico 5 – Interações totais nos recursos utilizados do AVEA: segundo experimento .....	59
Gráfico 6 – Média de utilização dos recursos no AVEA: segundo experimento ....	60
Gráfico 7 – Interações totais nos recursos utilizados do AVEA: terceiro experimento .....	65
Gráfico 8 – Média de utilização dos recursos no AVEA: terceiro experimento .....	66

## LISTA DE ABREVIATURAS E SIGLAS

ARFF	Attribute-Relation File Format
ABED	Associação Brasileira de Educação a Distância
EDM	Educational Data Mining
AVEA	Ambiente Virtual de Ensino e Aprendizagem
EBTT	Ensino Básico, Técnico e Tecnológico
IES	Instituições de Ensino Superior
IFFar	Instituto Federal de Educação, Ciência e Tecnologia Farroupilha
IFSul	Instituto Federal Sul-rio-grandense
KDD	Knowledge Discovery in Databases
MDE	Mineração de dados Educacionais
MEC	Ministério da Educação
SQL	Modular Object-Oriented Dynamic Learning
PRDI	Pró Reitoria de Desenvolvimento Institucional
UFPEL	Universidade Federal de Pelotas
UNIASSELVI	Universidade Luterana do Brasil
UNIVALI	Universidade do Vale do Itajaí
UFPR	Universidade de Brasília
UFRGS	Universidade Federal do Rio Grande do Sul
UFSC	Universidade Federal de Santa Catarina
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
TIC	Tecnologias de Informação e Comunicação
TCU	Tribunal de Contas da União

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>11</b>
1.1 PROBLEMA DE PESQUISA .....	14
1.2 OBJETIVOS .....	14
1.2.1 Objetivo geral .....	14
1.2.2 Objetivos específicos .....	14
1.3 JUSTIFICATIVA .....	15
1.4 ORGANIZAÇÃO DO TRABALHO .....	15
<b>2 REVISÃO DA LITERATURA</b> .....	<b>17</b>
2.1 A EDUCAÇÃO A DISTÂNCIA .....	17
2.2 EVASÃO NA EDUCAÇÃO A DISTÂNCIA .....	19
2.3 DADOS E CAUSAS DE EVASÃO NO BRASIL .....	21
2.4 A IMPORTÂNCIA DA TECNOLOGIA NA EAD .....	26
2.5 AMBIENTE VIRTUAL DE ENSINO-APRENDIZAGEM MOODLE – AVEA.....	27
<b>3 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS (KDD)</b> .....	<b>31</b>
3.1 PROCESSO DE KDD.....	31
3.1.1 Base de dados .....	32
3.1.2 Preparação ou pré-processamento de dados .....	32
3.1.3 Mineração de dados (MD).....	33
3.1.4 Avaliação de algoritmos .....	36
3.2 FERRAMENTA PARA MINERAÇÃO DE DADOS – WEKA .....	37
<b>4 TRABALHOS CORRELATOS</b> .....	<b>39</b>
4.1 DESCRIÇÃO DOS TRABALHOS.....	39
4.2 ANÁLISE DOS TRABALHOS CORRELATOS .....	41
<b>5 ASPECTOS METODOLÓGICOS</b> .....	<b>44</b>
6.1 PRIMEIRO EXPERIMENTO.....	46
6.1.1 Seleção dos dados.....	46
6.1.2 Preparação dos dados .....	48
6.1.3 Mineração de dados: primeiro experimento .....	51
6.1.4 Análise dos dados: primeiro experimento .....	55
6.2 SEGUNDO EXPERIMENTO .....	56
6.2.1 Mineração de dados: segundo experimento .....	56
6.2.2 Avaliação dos dados: segundo experimento.....	60
6.3 TERCEIRO EXPERIMENTO.....	61
6.3.1 Mineração de dados: terceiro experimento .....	61
6.3.2 Análise dos dados: terceiro experimento .....	66
6.4 ANÁLISE DA MINERAÇÃO ENTRE OS EXPERIMENTOS .....	68
<b>7 CONCLUSÃO</b> .....	<b>70</b>
<b>REFERÊNCIAS</b> .....	<b>73</b>
<b>ANEXO A – SOLICITAÇÃO</b> .....	<b>80</b>
<b>ANEXO B – AUTORIZAÇÃO</b> .....	<b>81</b>

## 1 INTRODUÇÃO

Há muito tempo discutem-se os problemas da evasão nos cursos de Educação a Distância (EaD) do Brasil, principalmente nas instituições públicas de ensino (TINTO, 1975). Com a disseminação das Tecnologias de Informação e Comunicação (TIC), o número de cursos EaD tende a ampliar, assim como seu acesso por toda organização e estudante.

Essa modalidade, já consolidada no Brasil, faz-nos ter outra perspectiva, devido ao grande número de matrículas efetivadas, todos os anos, em diferentes níveis de ensino. Esse destaque na educação fornece-nos dados que comprovam o tamanho da expansão – por exemplo, os 3.137 polos criados em 2017, sendo 30% deles em cidades onde as instituições sequer atuavam ainda (CENSO EAD, 2017).

Com esse avanço da oferta de polos EaD, o número de alunos também cresceu consideravelmente, sendo contabilizados 7.773.828 alunos (CENSO EAD, 2017). Tais dados mostram a capacidade e o crescimento da modalidade EaD, em diversas áreas do Brasil, potencializando ainda mais a expansão da educação.

Nesse contexto, surge a evasão, um tema tão discutido e fomentado em diversos contextos históricos, o qual devemos debater com maior cautela e tratar como assunto primordial no âmbito educacional. As taxas de evasão da EaD ainda são superiores às taxas dos cursos presenciais, porém esses dados estão cada vez mais próximos nas duas modalidades (CENSO EAD, 2017). Nesta pesquisa, existem algumas abordagens sobre a definição de evasão por diferentes autores. Dessa forma, para este estudo e para o IFFar, foi considerado evasão ou abandono o aluno que não se matriculou em, pelo menos, uma disciplina durante o semestre, ou seja, não manteve o vínculo com a instituição durante o período.

A evasão está presente em todas as modalidades de ensino, seja presencial, semipresencial ou a distância (BITTENCOURT; MERCADO, 2014). No contexto atual, há diversos trabalhos que buscam identificar as causas da evasão ou, até mesmo, que proponham metodologias e ferramentas para mitigá-las. Entre a gama de estudos, algumas pesquisas estão no entorno das Instituições de Ensino Superior (IES), porém este trabalho traz como foco a mineração de dados dos cursos técnicos subsequentes, ofertados pelo IFFar, por meio do Programa

Governmental Rede e-Tec.

A evasão escolar pode estar ligada a diversas causas, que estão diretamente relacionadas à qualidade da educação oferecida pela instituição de ensino frequentada pelo aluno, ao ambiente escolar, à relação familiar dele, ao meio social em que ele vive ou a motivos concernentes à vida pessoal do próprio aluno.

Nesse sentido, os motivos podem estar relacionados interna ou externamente à instituição, e, independentemente da categoria em que estejam inseridos, esses fatores precisam ser tratados pela gestão de EaD ou pelo órgão ao qual o aluno esteja ligado. Ações ou práticas pedagógicas são necessárias para incentivar o aluno, de forma que ele não se desestimule ou venha a evadir, prejudicando, assim, o aumento do índice de efetividade na educação e, conseqüentemente, um dos objetivos das IES, a permanência e o êxito.

Nesse processo pedagógico, no qual temos a EaD como uma alternativa do processo de ensino-aprendizagem, as TIC surgem como uma das principais ferramentas de sustentação dessa modalidade de ensino, em constante expansão.

Nesse contexto, estão inseridos os Ambientes Virtuais de Ensino-Aprendizagem (AVEA), responsáveis pela interação entre aluno, professor e tutor, sendo possível compartilhar materiais, realizar tarefas, interagir com outros alunos, participar de fóruns e *Wikis*, bem como outras atividades inerentes ao processo de ensino e aprendizagem EaD.

Entretanto, entre as diversas interações entre aluno e professor pela ferramenta, ainda se encontram dificuldades para acompanhar o desempenho do estudante em cursos oferecidos de forma virtual, ou seja, pelo AVEA. Conforme Oliveira et al. (2012), os índices de reprovação, evasão e desistência são relativamente altos e, sendo assim, diferentes pesquisas são, cada vez mais, necessárias para definirmos as causas da evasão e formas de acompanhar o desenvolvimento do aluno. Assim, é importante predizer a sua evasão dentro do ambiente.

No AVEA, em forma de registro de *logs*, há inúmeras informações brutas que podem ser exploradas para auxiliar o gestor na tomada de decisão, a fim de mitigar a evasão. As tecnologias devem ser usadas sempre em prol da educação, pois, assim como permitem a oferta de cursos EaD, podem e devem ser utilizadas para maximizar a permanência do aluno no curso. Para isso, é necessária a criação de métodos e meios que auxiliem nesse processo minucioso.

Verificar as causas da evasão, acompanhar o andamento do aluno, diagnosticar com antecedência as dificuldades de aprendizagem, são formas de auxiliar o êxito na educação e subsidiar os gestores e professores das instituições para práticas pedagógicas mais eficientes.

De modo que possa auxiliar nesse tema e iniciar o processo de reduzir a evasão por algum estudo, o Instituto Federal Farroupilha (IFFar), por meio da Resolução 178/2014, criou o Programa Permanência e Êxito, em 2016, com o objetivo de consolidar a excelência da oferta do Ensino Básico, Técnico e Tecnológico (EBTT) de qualidade e promover ações para a permanência e o êxito dos estudantes no IFFar. Esse programa relata causas e fatores externos, ou seja, estudos feitos após o aluno sair da instituição, por hipóteses e relatos de experiência dos Coordenadores de Registros Acadêmicos (CRA).

Desse modo, para que a tecnologia possa auxiliar na descoberta de conhecimento e verificar os fatores que têm influenciado a evasão, com base nos registros de dados dos alunos no sistema, faz-se necessário recorrermos a uma técnica chamada Mineração de Dados (MD). O grande número de dados obtidos hoje, em virtude do número elevado de alunos presentes nos AVEA, traz ainda mais destaque para a MD.

Com o propósito de buscar medidas para reduzir a evasão, a mineração surge como excelente alternativa, pois possibilita identificar quais os fatores que contribuem para o insucesso dos estudantes dentro dos ambientes virtuais de aprendizagem. De forma a entendermos melhor sobre a definição de MD, podemos dizer que ela consiste na exploração de grandes quantidades de dados com o objetivo de detectar padrões que permitem a extração de novos conhecimentos (SILBERSCHATZ; KORTH; SUDARSHAN, 2006). Na literatura, a MD, em ambientes educacionais, é tratada como Mineração de Dados Educacionais (MDE), do inglês *Educational Data Mining* (EDM).

Portanto, a partir da constante expansão do ensino a distância e da necessidade de potencializarmos esse ensino com o uso das tecnologias, aliado a diversos estudos, a mineração de dados torna-se grande aliada da modalidade. Seu emprego em um grande número de aplicações científicas mostra-se viável para resolver diversos problemas relacionados à investigação de informações úteis em bases de dados.



## 1.1 PROBLEMA DE PESQUISA

A permanência e o êxito é um assunto bastante abordado nas IES, com diferentes debates e reflexões acerca da evasão. Desse modo, a instituição de ensino e os gestores educacionais precisam de dados e informações qualificadas capazes de identificar os motivos da evasão no contexto da educação.

O abandono do aluno representa um prejuízo para diferentes partes, seja para a instituição, seja para o próprio aluno. No entanto, as causas da evasão, apesar de gerarem grande impacto para a educação, não são amplamente conhecidas pelas IES, conforme mostra o Censo da EaD, realizado pela Associação Brasileira de Educação a Distância (CENSO EAD, 2017). Entre os cursos regulamentados totalmente a distância, 59% das instituições respondem que não sabem os reais motivos de abandono.

## 1.2 OBJETIVOS

Os objetivos deste trabalho estão divididos em objetivo geral e objetivos específicos.

### 1.2.1 Objetivo geral

Este estudo tem como objetivo geral, por meio da Mineração de Dados Educacionais (MDE), analisar, por meio da interação dos alunos no AVEA, possíveis evasões em cursos do Instituto Federal Farroupilha na modalidade a distância, disponibilizando dados estratégicos para os gestores educacionais da instituição.

### 1.2.2 Objetivos específicos

Buscando-se atingir o objetivo geral desta pesquisa, destacam-se os seguintes objetivos específicos:

- a) Realizar busca de dados de evasão no IFFar;
- b) Revisar na literatura as principais causas de evasão educacional;
- c) Identificar as informações disponíveis no AVEA que podem prever a evasão;

- d) Avaliar técnicas de MDE que auxiliem na análise preditiva de evasão;
- e) Realizar a mineração de dados educacionais em turmas subsequentes da EaD;
- f) Avaliar qualitativamente os dados disponibilizados pela mineração de forma a verificar a sua contribuição para os gestores do IFFar.

### 1.3 JUSTIFICATIVA

Com a expansão da EaD no IFFar cada vez mais apoiada nas tecnologias educacionais, surge a necessidade de direcionamento de ações em nível institucional, sobretudo esforços para aumentar o número de alunos matriculados nos cursos ofertados pela Instituição.

Nesse contexto, este projeto propõe a mineração de dados como forma de buscar padrões de perfis de acesso ao AVEA do IFFar.

Conforme Plano de Desenvolvimento Institucional (PDI 2019-2026) do IFFar, algumas metas e estratégias para a EaD são:

- Ampliar a oferta de cursos na modalidade de Educação a Distância (EaD);
- Reduzir o índice de evasão dos cursos EaD de 36,3% para 23%, até 2026;
- Buscar alternativas para melhoria do desempenho dos estudantes, visando prevenir e melhorar a evasão e a retenção escolar.

Sendo assim, temos a tecnologia como grande ferramenta de apoio aos gestores educacionais das IES. Com os avanços tecnológicos, os AVEA possibilitam a busca de novos conhecimentos dentro da base de dados pelo processo de Knowledge Discovery in Databases (KDD) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

### 1.4 ORGANIZAÇÃO DO TRABALHO

O restante deste trabalho está organizado da seguinte forma: o capítulo 2 apresenta uma fundamentação teórica do estudo e está estruturado nas seguintes seções: “A educação a distância”; “A evasão na educação a distância”; “Dados e causas da evasão no Brasil”; “O papel tecnológico do gestor na evasão” e, por fim, “Ambiente virtual de aprendizagem Moodle”.

No terceiro capítulo, são apresentados os conceitos de KDD e MDE. No

quarto capítulo, os trabalhos correlatos ao desenvolvido nesta pesquisa. Na sequência, no capítulo cinco, os aspectos metodológicos. O desenvolvimento do processo de mineração de dados é apresentado no capítulo sexto. O sétimo traz a conclusão. Por fim, o oitavo, as referências bibliográficas.

## 2 REVISÃO DA LITERATURA

Este capítulo aborda os conteúdos e conceitos bases para a realização desta pesquisa. Primeiramente, é abordada a definição da EaD, sua contextualização e suas características. Posteriormente, serão apresentados os dados e as causas da evasão no Brasil. Na sequência, é abordado brevemente o papel do gestor na evasão dentro da IES, como forma de apontarmos a sua importância e também a tecnologia no contexto da evasão, e, por fim, o AVEA.

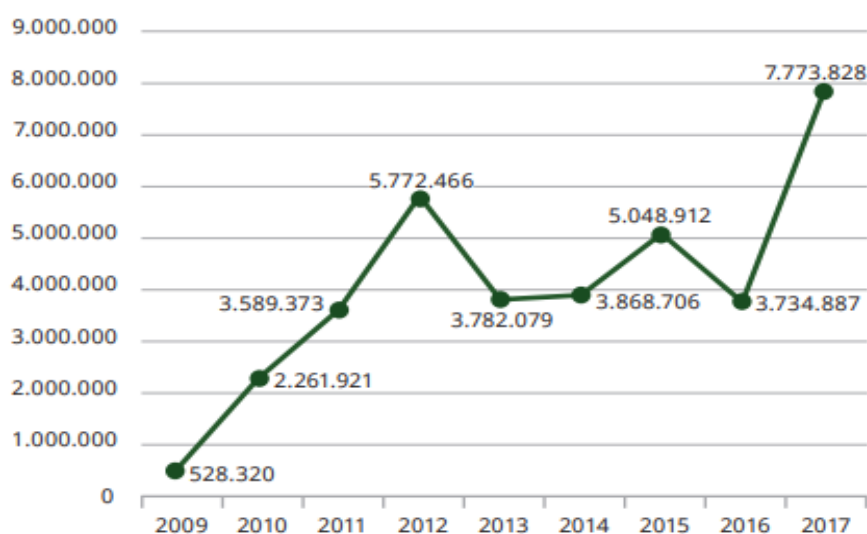
### 2.1 A EDUCAÇÃO A DISTÂNCIA

A EaD é uma modalidade de ensino em que alunos e professores encontram-se em espaços distintos, podendo-se definir, segundo Moran (2002), que a educação a distância é o processo de ensino-aprendizagem mediado por tecnologias, no qual professores e alunos estão separados espacial e/ou temporalmente. Isto é, a EaD tem sido considerada uma forma complementar para a formação do cidadão e tem se mostrado bastante rica em potenciais pedagógicos e de democratização do conhecimento (MILL, 2013). Nesse sentido, a tecnologia é fundamental para que a relação entre aluno e professor aconteça de forma dinâmica e eficiente, apropriando-se de diferentes recursos tecnológicos e inovadores, proporcionando aos seus alunos um ambiente capaz de buscar a evolução e o aprendizado nos diferentes espaços educacionais ofertados por essa modalidade de ensino.

No âmbito do estudante, nota-se que o ensino a distância proporciona acesso à capacitação e formação de indivíduos que estão em locais distantes dos grandes centros de ensino, bem como permite uma maior flexibilidade e autonomia em relação ao estudo, sem perder a qualidade no ensino. Vale ressaltar que cabe ao aluno administrar seu próprio aprendizado e ter responsabilidades em estabelecer seus próprios objetivos quanto ao seu estudo (BELLONI, 2001).

Como forma de evidenciar a evolução da educação a distância desde 2009, a seguir, apresenta-se a série histórica com o volume total de matrículas contabilizadas pelo CENSO de 2017.

Gráfico 1 – Número de matrículas na EaD ao longo do tempo

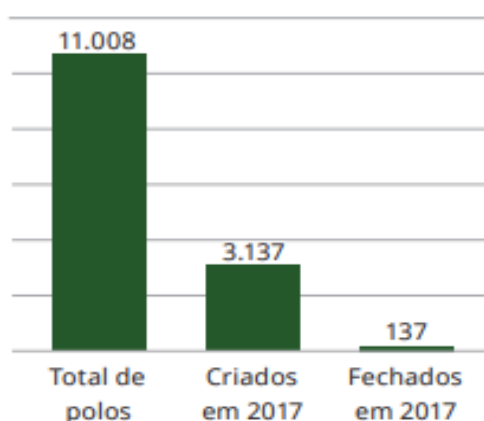


Fonte: ABED, 2017.

No gráfico 1, pode-se perceber um crescimento significativo desde 2009 na educação a distância, o que indica a necessidade e a devida importância nessa modalidade para o crescimento da educação no Brasil.

Nesse avanço, o Censo 2017 contabilizou um crescimento no número de polos existentes – do total de 11.008 polos contabilizados, 3.137 foram criados em 2017. Dessa forma, evidencia-se que a expansão da educação, bem como da rede de instituições, tem avançado sobre as mais diversas cidades. Esses dados podem ser mais bem visualizados no gráfico 2:

Gráfico 2 – Polos EaD



Fonte: ABED, 2017.

Pesquisas da Associação Brasileira de Educação a Distância (2016) indicam os principais problemas enfrentados pela EaD no Brasil: o principal é a evasão, bastante abordada neste estudo, caracterizada pela desistência do estudante do curso; o segundo é a resistência dos docentes em relação à modalidade, muitas vezes, causada justamente pela falta de conhecimento sobre TIC e sobre o devido domínio nas plataformas EaD; por fim, a dificuldade de adaptação dos estudantes à modalidade EaD. Esses problemas podem causar certa frustração aos alunos devido à metodologia de ensino ser diferente da abordada na modalidade presencial, pois deixamos de ter a figura do professor, acabando por, até mesmo, dar equivocadamente menor importância ao ensino a distância.

A seguir, abordaremos o número de evasões na EaD e seus conceitos.

## 2.2 EVASÃO NA EDUCAÇÃO A DISTÂNCIA

A evasão pode estar ligada a diferentes motivos particulares de cada IES e também a cada necessidade não correspondida para as diferentes realidades dos alunos inseridos nesse campo educacional. A fim de entender os conceitos de evasão, discutimos, a seguir, algumas definições e classificações de alguns autores.

A evasão, que se torna um dos maiores desafios enfrentados pela EaD, é definida, segundo Maia (2007), como a desistência do aluno em completar o curso, independentemente se cursou aulas ou não, ou seja, aquele que desiste definitivamente do curso em qualquer etapa (FAVERO, 2006; ABBAD; CARVALHO; ZERBINI, 2006).

Outros autores como Toczec et al. (2008) a definem como o desligamento ou abandono do aluno da instituição de ensino, que pode ser compreendido como um processo individual, mas também pode constituir-se em coletivo.

Para Eyng et al. (2013), a evasão é definida como um processo de abandono da escola no ano letivo, levando o estudante a deixar de frequentar as aulas, bem como de realizar suas atividades curriculares, ocorrendo de forma gradativa no andamento do curso, até mesmo sem que isso seja notado pelos gestores escolares. Percebe-se que isso ocorre no decorrer das aulas do curso, sem ser observado de forma a evitar que o abandono do aluno ocorra, destacando ainda mais o papel primordial do professor para mitigar a evasão na educação, um dos

principais desafios da EAD.

Martínez, García e Montoro (2003) ainda classificam a evasão de quatro formas:

- a) Evasão: quando o aluno abandona o curso durante o seu período sem retornar;
- b) Trancamento: interrupção temporária do curso;
- c) Evasão do curso: o aluno abandona o curso antes do término, porém obtém conhecimento;
- d) Não iniciado: o aluno não inicia o curso.

Já para Manhães et al. (2011), a evasão pode ser definida em três eixos:

- a) A evasão de curso: caracterizada por abandono do estudante, desistência, transferência ou, até mesmo, pela sua exclusão do curso por norma institucional;
- b) Evasão da instituição: o estudante desliga-se da instituição na qual está matriculado;
- c) Evasão do sistema: abandono definitivo ou temporário.

A evasão ou abandono escolar é definida como um processo que tem natureza multiforme: a escolha de sair da escola é apenas o ato final de um processo que se manifesta de muitas maneiras, visíveis ou não, ao longo da trajetória escolar do indivíduo. Devem-se, portanto, detectar os sinais enviados pelos alunos quando estão em situação de risco (DORE; LÜSCHER, 2011).

Segundo Manhães et al. 2011, para reduzir o problema da evasão, é primordial a detecção dos alunos com tendência a evadir, de modo que possamos atentar especificamente ao aluno ou grupo de alunos, dispondo de atendimento diferenciado para tratar o problema. Segundo o CENSO EAD 2017, é surpreendente o fato de que menos de 50% das instituições conheçam os motivos dessa evasão.

Como forma de buscar sempre a eficiência de uma instituição de ensino, os gestores, cada vez mais, buscam ferramentas e métodos que o auxiliem a mitigar a evasão do aluno. Essas instituições têm apresentado, muitas vezes, altos índices de evasão escolar, seja em cursos técnicos ou superiores da instituição de ensino. Nesse contexto, deve-se sempre buscar indicadores de evasão e os motivos a qual leva o aluno a evadir do curso matriculado, já que esta informação pode auxiliar e muito os gestores na tomada de decisão (OLIVEIRA et al., 2017).

O ensino a distância precisa de um maior cuidado e atenção se comparado

com a modalidade presencial, justamente pelo fato de a figura do professor não estar frente ao aluno no seu cotidiano. Este tem o controle do seu tempo, faz suas atividades quando pertinente, devido à flexibilidade que a modalidade EAD disponibiliza. As tarefas curriculares passam a concorrer com as tarefas pessoais de forma a exigir uma maior organização pessoal e concentrações para o estudo (BITTENCOURT; MERCADO, 2014).

A forma de estudar, as características e o perfil do estudante da educação virtual, efetivada no ciberespaço, bem como suas estratégias de organizar seus estudos, mudam sobremaneira e, por isso, é importante que educadores e gestores conheçam o estudante virtual (MILL, 2018).

### 2.3 DADOS E CAUSAS DE EVASÃO NO BRASIL

De acordo com os censos anuais, as taxas de evasão são maiores nos cursos a distância quando comparados com os cursos presenciais. Segundo o Censo EAD.BR 2016, elaborado pela Associação Brasileira de Educação a Distância (ABED, 2017), há uma taxa de evasão entre 11% e 25% nas instituições com cursos totalmente a distância.

Woodley e Simpson (2015), ao realizarem uma busca no Google Acadêmico, em 18 de maio de 2018, com a expressão *allintitle*: evasão “educação a distância” OR EaD, encontraram 86 resultados. A mesma busca foi realizada pelo autor dessa dissertação em 24 de abril de 2019, com retorno de 146 resultados. Isso nos mostra uma evolução nas pesquisas e maior preocupação com o tema, além da necessidade de maiores esforços dedicados a mudar as taxas de retenção no ensino a distância.

De modo a diminuir essas taxas, precisamos primeiramente identificar as causas que levam o aluno a evadir. Entre os fatores, encontram-se a falta da tradicional relação entre aluno e professor, o insuficiente domínio do uso do computador por parte do aluno, a dificuldade do aluno em expor ideia em uma comunicação escrita a distância, o cansaço ao final do dia de trabalho, a ausência de tempo e de condições financeiras (COELHO, 2002; FAVERO, 2006; SANCHEZ, 2008).

Entre os motivos investigados pelas instituições e declarados no Censo EAD.BR 2016, a questão financeira é apontada como principal causa de evasão,



juntamente com a falta de tempo e a falta de adaptação à modalidade. Por isso, as IES precisam criar estratégias de suporte ao aluno, para, então, poder ajudá-lo a encontrar soluções para aquilo que se torna um empecilho no meio do seu processo educacional.

Um fato importante, ainda, é que os adultos são a maioria do público nas IES EAD, a maior parte dos estudantes está nas faixas etárias entre 26 e 30 anos e 31 a 40 anos. Então, pode-se concluir que a maioria desse público não cresceu com tantos avanços tecnológicos e, por isso, ainda não se adaptou ou não possui tanta facilidade para lidar com as plataformas educacionais ofertadas quanto os adolescentes, por exemplo (CENSO EAD, 2017).

Para Barroso e Falcão (2004), os fatores em destaque que podem desencadear a evasão escolar são:

- a) Econômicos: impossibilidade de permanecer no curso por questões socioeconômicas;
- b) Vocacionais: o aluno não se identifica com o curso; e
- c) Institucionais: abandono por fracasso nas disciplinas iniciais, inadequação aos métodos de estudo, dificuldades de relacionamento com colegas ou com membros da instituição.

Já para Lobo (2012), em seus estudos realizados, as causas mais encontradas são:

- a) Inadaptação do aluno ao estilo do Ensino Superior e falta de maturidade;
- b) Formação básica deficiente;
- c) Dificuldade financeira;
- d) Irritação com a precariedade dos serviços oferecidos pela IES;
- e) Decepção com a pouca motivação e atenção dos professores;
- f) Dificuldades com transporte, alimentação e ambientação na IES;
- g) Mudança de curso.

Ramos (2014), em uma pesquisa realizada no período de 2007 a 2011, sobre evasão e permanência, destaca que esse assunto pode ser direcionado pelos seguintes focos: estudantes, instituições, administradores e corpo docente. Ainda, aborda duas outras variáveis que podem levar à evasão, classificando-as como fatores anteriores e posteriores à admissão do curso. Os fatores anteriores ou pré-admissão no curso são as características dos alunos (idade, etnia, gênero,

desenvolvimento intelectual e desempenho), competências e habilidades (letramento digital, letramento informacional, administração do tempo, leitura e escrita, interação com computador).

Já os fatores posteriores ou pós-admissão podem ser divididos, segundo Ramos (2014), em fatores internos e externos, mais bem ilustrados no quadro 1:

Quadro 1 – Fatores internos e externos da evasão

<b>Fatores internos</b>
Finanças
Horas de trabalho
Responsabilidades familiares
Falta de incentivo
Crises da vida
<b>Fatores externos</b>
Integração escolar e social
Comprometimento com metas e objetivos
Comunidade de aprendizagem
Clareza do programa
Autoestima
Relações interpessoais
Acessibilidade aos serviços
Hábitos de estudo
Orientação
Absentismo
Ajuste do programa
Stress
Satisfação
Comprometimento
Estilo de aprendizagem e ensino

Fonte: Ramos (2014, adaptado).

Esses diferentes fatores são mais bem compreendidos por Ramos, Bicalho e Sousa (2015), que apontam a evasão ligada ao comportamento de pessoas, em que a decisão de abandonar o curso é do estudante, podendo as causas e a origem da evasão estarem ou não sob o conhecimento da instituição, podendo ser ainda composta dos mais variados fatores.

Nesse contexto, Pereira (2003, p. 54) afirma que a evasão escolar pode estar ligada a diversos fatores e não a um especificamente, em que a melhor forma é apurá-los e tratá-los para que o aluno retorne à escola.

Johann (2012, p. 66) relata que a evasão é um fenômeno complexo, podendo ocasionar mudanças sociais e econômicas, pois também pode trazer perdas

financeiras para as instituições. Essas perdas geram uma série de consequências dentro da escola e, até mesmo, para o governo.

Bittencourt e Mercado (2014), ao buscarem as causas da evasão, evidenciaram o que já foi abordado aqui pelos outros autores: que a evasão está diretamente ligada a fatores internos e externos, classificados como endógenos e exógenos, ou seja, fatores institucionais e didático-pedagógicos e fatores sobre os quais a instituição não tem o controle e a possibilidade de intervir junto ao aluno, respectivamente.

Bizarria, Silva e Carneiro (2014) apontam o papel do tutor em uma instituição pública, em que o tutor com uma abordagem pedagógica tem mais chances de constituir laços com o aluno. Isso, juntamente com a interação, tornam-se fatores determinantes para reduzirmos a evasão.

De modo a enriquecer ainda mais este estudo, uma pesquisa exploratória é realizada por Fiuza (2012), a partir de um questionário *online* com 605 estudantes de cursos de graduação e pós-graduação de universidades públicas brasileiras – Universidade Federal de Santa Catarina (UFSC), Universidade Federal do Paraná (UFPR), Universidade Federal do Rio Grande do Sul (UFRGS) e Universidade de Brasília (UnB) – e três instituições privadas – Centro Universitário Leonardo da Vinci (UNIASSELVI), Universidade do Vale do Itajaí (UNIVALI) e Universidade Luterana do Brasil (ULBRA).

A tabela 1 apresenta os resultados da abordagem sobre as principais dificuldades ou problemas encontrados pelos discentes para permanecerem no curso, conforme estudo realizado por Fiuza (2012).

Tabela 1 – Incidência das categorias de atividades que contribuíram para a permanência

(continua)

<b>Categoria</b>	<b>Número de respostas na categoria</b>	<b>% Categoria</b>
Atitude/Comportamento do professor	78	18,10
Atividades	72	16,71
Motivação/Incentivos	56	12,99
Atitude/Comportamento do tutor	49	11,37
Material didático/Conteúdo	27	6,26
Questões afetivas/sentimentais	25	5,80
Aulas	24	5,57
Recursos tecnológicos	24	5,57
Interação/Comunicação	23	5,34
Persistência	16	3,71

Tabela 1 – Incidência das categorias de atividades que contribuíram para a permanência

		(conclusão)
Disciplinas	16	3,71
Desempenho pessoal	11	2,55
Flexibilidade de horários	10	2,32
<b>Total</b>	<b>431</b>	<b>100</b>

Fonte: Fiuza (2012, p. 85).

Os dados mostram a relevância das categorias, bem como os motivos de permanência, que estão diretamente ligados aos professores que atuam na EaD e às suas ações.

Fiuza (2012) ainda relaciona os motivos que levam à evasão, conforme o quadro 2.

Quadro 2 – Motivos da evasão EaD

<b>Motivos que levam à evasão</b>
Qualidade da aula
Fatores socioeconômicos
Apatia
Vida pessoal/familiar
Trabalho
Problemas com a tecnologia
Dificuldade de acesso a um computador
Velocidade de conexão lenta
Falta de autonomia do estudante
Falta de apoio acadêmico/administrativo
Resistência com a tecnologia
Renda familiar baixa

Fonte: Elaborado a partir de Fiuza (2012).

Outro ponto importante que cabe destacar aqui é a localização dos polos presenciais de Educação a Distância. Estes estão situados, muitas vezes, em municípios do interior, explica Andrade (2010). Isso acaba trazendo certo problema dependendo da região, pois a infraestrutura é fundamental para a modalidade EaD e utilização dos alunos, podendo-se citar como exemplo crítico a tecnologia e a internet.

Dessa forma, as dificuldades de distância e, conseqüentemente, o tempo de deslocamento que os alunos sofrem para chegar até os polos de apoio presencial podem também levar à evasão (ANDRADE, 2010).

Sendo assim, pode-se perceber, ao longo dessas discussões e da forma como os autores abordam os diferentes motivos da evasão nos cursos EAD, que é preciso uma busca constante em aliar as tecnologias presentes no ensino àquilo que o aluno se propôs a buscar dentro da IES na qual ele está inserido. Assim também, é preciso buscar novas formas de aprender e se relacionar com a plataforma, com os professores, com os colegas e com a instituição à qual ele faz parte (OLIVEIRA, 2009; MAIA; MATTAR, 2007).

Portanto, há uma grande variedade de motivos que podem estar ligados à evasão dos alunos no EaD, de maneira direta ou indireta. As IES precisam se conscientizar de que as tecnologias são grandes aliadas para auxiliar nesse processo de busca e aperfeiçoamento daquilo que pode se tornar um grande problema, a evasão EaD.

Na próxima seção, será abordado brevemente o papel tecnológico do gestor, bem como sua importância nesse contexto.

## 2.4 A IMPORTÂNCIA DA TECNOLOGIA NA EAD

Compreendendo os aspectos históricos, culturais e tecnológicos que envolvem os processos educacionais na contemporaneidade, Mill (2013, p. 11-12) ressalta que:

[...] a cultura educacional está diretamente relacionada com as tecnologias disponíveis, e o uso que se faz destas no campo educacional relaciona-se, portanto, às suas potencialidades pedagógicas. Nesse sentido, é importante em seus quatro elementos constitutivos, isto é: gestão (gestores), ensino (educadores), aprendizagem (estudantes) e mediação tecno-pedagógica (tecnologias). Entendemos que a educação, em qualquer época ou lugar, constitui-se e fundamenta-se nesses quatro elementos de modo articulado, complementar, dinâmico e dialético. Portanto, a análise da incorporação das tecnologias digitais no âmbito educacional deve ser feita considerando os outros três elementos (gestão/ensino/aprendizagem).

Nesse contexto, esses elementos reforçam ainda mais a necessidade de educadores e gestores que utilizem o máximo das tecnologias para melhorar a educação a distância.

Em uma pesquisa realizada pelo CENSO EAD 2017, identificou-se que as estratégias de gestão são do interesse de apenas 53% dos respondentes. O CENSO 2017 ainda afirma que os seis elementos que adquiriram o maior grau de

associação à qualidade na EAD independem da modalidade, pois se associam à educação em geral. Os elementos são: “conteúdos corretos e atualizados”, “professores qualificados”, “tutores qualificados”, “atendimento ágil às necessidades dos alunos”, “metodologias eficazes” e “gestão eficaz”.

Nesse contexto, a utilização das tecnologias no campo educacional e o aumento do uso dos ambientes virtuais fazem com que as instituições e os gestores tenham grande quantidade de dados envolvendo estudantes, tutores e professores. Rigo et al. (2012) destacam que esse alto número de dados tem sido utilizado de forma ineficiente para colaborar com a monitoração, predição e mitigação da evasão escolar, sendo impossível tratarmos os dados manualmente. Devido a esse avanço tecnológico, está cada vez mais difícil gerenciar dados, sendo um desafio para as instituições de ensino, conforme destacam Romero e Ventura (2013).

As tecnologias podem ser usadas de forma integradora no processo de aprendizagem na EaD, enfatiza Mill (2010), independentemente das pessoas que hoje a consideram o futuro, a salvação, e até mesmo aquelas que não são adeptas a esse fim. No entanto, devido ao grande número de alunos matriculados nos cursos EaD, sob responsabilidade dos gestores, fica inviável conhecer as causas da evasão de todos os estudantes. Nesse contexto, os AVEA são ferramentas importantes e podem contribuir para obtenção de dados que possam identificar a evasão antecipada desses alunos.

Sendo assim, a próxima seção abordará o conceito e as características de AVEA.

## 2.5 AMBIENTE VIRTUAL DE ENSINO-APRENDIZAGEM MOODLE – AVEA

Com o crescimento das Tecnologias de Informação e Comunicação (TIC), o ensino a distância também evolui à medida que novas ferramentas surgem para auxiliar no processo de ensino aprendizagem. Entre os recursos existentes, destacam-se os Ambientes Virtuais de Ensino-Aprendizagem (AVEA).

Nesse rol, pode-se citar os mais utilizados no âmbito nacional, a exemplo: Moodle, Teledu e E-Proinfo. No entanto, o Moodle é o mais conhecido mundialmente, por ser um *software* livre e gratuito. Nesse contexto, há diversos usuários com o propósito de aperfeiçoar o desenvolvimento da ferramenta educacional.

Segundo Almeida (2003, p. 311), AVEA são:

sistemas computacionais disponíveis na Internet, destinados ao suporte de atividades mediadas pelas tecnologias de informação e comunicação. Permitem integrar múltiplas mídias, linguagens e recursos, apresentar informações de maneira organizada, desenvolver interações entre pessoas e objetos de conhecimento, elaborar e socializar produções tendo em vista atingir determinados objetivos.

Para atuar na área da educação, hoje, faz-se necessário estar sempre inovando e buscando novos métodos para ensinar/aprender. Nos ambientes virtuais, esse processo de ensinar e aprender, por algumas vezes, não se torna uniforme, consciente por ambas as partes. Conforme Moran (2007, p. 2), educação *online* é o local onde:

[...] o aluno se conecta a uma plataforma virtual e lá encontra materiais, tutoria e colegas para aprender com diferentes formas de organização da aprendizagem: umas mais focadas em conteúdos prontos e atividades até chegarmos a outras mais focadas em pesquisa, projetos e atividades colaborativas, onde há alguns conteúdos, mas o centro é o desenvolvimento de uma aprendizagem ativa e compartilhada.

Entre as diversas plataformas existentes, para que o ambiente virtual seja de aprendizagem, são necessários os seguintes requisitos, conforme indica (SANTOS, 2003, p. 227):

- a) Criar sites hipertextuais que agreguem intertextualidade, conexões com outros sites ou documentos;
- b) Potencializar comunicação interativa síncrona, comunicação em tempo real e assíncrona, comunicação a qualquer tempo;
- c) Criar atividades de pesquisa que estimulem a construção do conhecimento a partir de situações problemas, em que o sujeito possa contextualizar questões locais e globais do seu universo cultural;
- d) Criar ambiências para avaliação formativa, em que os saberes sejam construídos num processo comunicativo de negociações e a tomada de decisões seja uma prática constante para a (re)significação processual das autorias e coautorias;
- e) Disponibilizar e incentivar conexões lúdicas, artísticas e navegações fluidas.

Portanto, a educação *online* é potencializada pelos ambientes virtuais de aprendizagem (AVA), ou seja, as chamadas plataformas de colaboração. Conforme Kenski (2007, p. 94), os AVA são “sistemas computacionais disponíveis na Internet, destinados ao suporte de atividades mediadas pelas tecnologias de informação e comunicação”.

Nesse sentido, o ambiente virtual Moodle é considerado a plataforma de aprendizagem mais popular do mundo, com mais de 142 milhões de cadastros registrados, conforme o site (<https://moodle.net/stats/>), e utilizado em larga escala nas escolas, universidades e empresas. Sua popularidade deve-se ao fato de ser flexível e personalizável, ou seja, é uma plataforma de código aberto. Além disso, o Moodle oferece diversas ferramentas interessantes para o ensino, como fóruns, wikis, tarefas, enquetes e a possibilidade de o usuário inserir outros recursos interativos, como arquivos de áudio e vídeo, enviar *feedback* a outros usuários, Webconference Internet (conferências com imagens e som), Vídeo-Tutor (conjugação de páginas HTML com vídeo), *Mobile Learning Engine* (dispositivos móveis) e *Mobile QUIZ* (questionários em forma de testes para serem respondidos por dispositivos móveis) (SILVA, 2010).

Nesta dissertação, aborda-se como foco o ambiente Moodle utilizado no IFFar para o ensino a distância.

A figura 1 mostra a interface do Moodle utilizada no Instituto Federal Farroupilha.



Figura 1 – Interface Moodle no IFFar

Fonte: <<https://MOODLE2.iffarroupilhaead.edu.br/>>.

Além das ferramentas e dos recursos educacionais aqui citados, o Moodle também fornece algumas ferramentas de gestão, por exemplo, a ferramenta relatórios. Os relatórios permitem a visualização geral das ações do usuário no ambiente, tais como os acessos dos alunos no AVEA, a realização das tarefas, a participação em fóruns e outras informações importantes para o acompanhamento do aluno (SILVA, 2011). Todas as interações realizadas pelos alunos no AVEA são armazenadas na base de dados do Moodle, conhecida como banco de dados (BD).

Essas informações possuem grande importância para a Mineração de Dados (MD), a qual será abordada na próxima seção deste trabalho, que consiste em uma técnica para a descoberta de dados implícitos, podendo ou não ser estratégicos para a gestão da instituição.

### 3 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS (KDD)

Atualmente, com o avanço da tecnologia e com o crescimento dos sistemas de informações, encontram-se grandes bases de dados em diversas organizações. O armazenamento está em diversas áreas, seja em base de natureza administrativa, comercial, educacional, governamental ou social. A análise desse grande volume de dados requer o uso de tecnologias que auxiliem no processo, pois, devido à quantidade de informações, torna-se inviável para o homem analisá-los sem o auxílio das ferramentas apropriadas.

Entretanto, precisa-se utilizar ferramentas que possam extrair esses dados em forma de conhecimento e, assim, analisar, interpretar e relacioná-los para a tomada de decisão, de modo que estratégias sejam criadas juntamente com a otimização de processos gerenciais (GOLDSCHMIDT; BEZERRA; PASSOS, 2015).

Nesse sentido, surge a Descoberta de Conhecimento em Base de Dados (do inglês *Knowledge Discovery in Databases* – KDD), em que uma das etapas é a Mineração de Dados.

KDD é definido por Fayyad, Piatetsky-Shapiro e Smyth (1996) como “o processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. Em suma, o KDD é empregado na identificação de padrões por meio da manipulação de dados (AGRAWAL et al., 1993; BRACHMAN et al., 1996; MITCHELL, 1999).

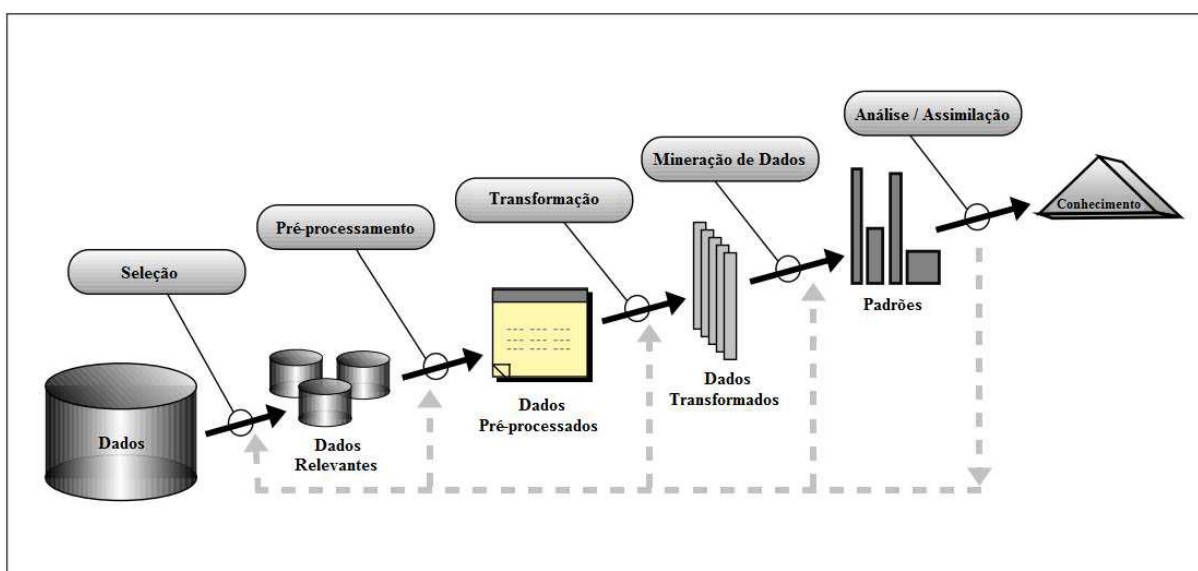
Além disso, devemos avaliar os critérios de restrição ou expectativa dos algoritmos quanto a sua precisão, por meio da aplicação das técnicas de MD. Essa avaliação está diretamente relacionada à qualidade e quantidade de dados, abordadas no item 3.1.3.

#### 3.1 PROCESSO DE KDD

O processo KDD é composto por várias etapas, que vão desde a seleção e amostragem dos dados, limpeza, enriquecimento e preparação, até a mineração de dados e avaliação. Essas etapas serão detalhadas nas próximas seções. Na figura

2, baseada nas considerações de Fayyad, Piatetsky-Shapiro e Smyth (1996), são apresentadas todas as etapas envolvidas no processo de KDD.

Figura 2 – Etapas do processo de KDD



Fonte: Adaptada de Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 29).

Para um melhor entendimento, nas próximas subseções, serão caracterizados os itens que compõem o processo.

### 3.1.1 Base de dados

As bases de dados utilizadas no processo de KDD podem ser obtidas de diversas formas de estruturas e armazenamento, tais como: bancos de dados relacionais, planilhas eletrônicas, *data warehouse*, arquivos de *log*, *data stream*, dados da WEB, entre outros (MANHÃES, 2015). A limpeza e transformação de dados são passos necessários para melhorar sua qualidade, a fim de, posteriormente, aplicarmos as técnicas de MD.

### 3.1.2 Preparação ou pré-processamento de dados

Como fase que antecede a mineração, é necessário efetuar a seleção dos dados considerados importantes para a organização (CASTRO; FERRARI, 2016). As

funções básicas de pré-processamento dos dados serão descritas a seguir.

1) Limpeza dos dados: para que possamos assegurar a qualidade, veracidade e integridade dos dados, é necessário realizarmos uma limpeza nestes, ou seja, eliminar inconsistências e, até mesmo, dados desnecessários, para que, no final, não se comprometa o processo de KDD. Alguns exemplos de operações de limpeza são: eliminação de dados errôneos, padronização de dados, eliminação de dados duplicados (HAN; KAMBER, 2001).

2) Seleção: possui como finalidade buscar os dados relevantes à análise (HAN; KAMBER, 2001). Diante da grande quantidade de dados disponíveis, é preciso selecionar as tabelas e atributos mais utilizados como forma de não utilizarmos os dados irrelevantes na MD.

3) Transformação: muitas vezes, precisamos converter os dados para o formato apropriado à mineração, existindo, assim, a aplicação de operações de transformação na fase de pré-processamento. Alguns exemplos de transformação de dados são: conversão de valores simbólicos para numéricos, normalização de dados (HAN; KAMBER, 2001). A seguir, serão abordados dois tipos de transformações, a padronização e a discretização:

a) *Padronização*: objetiva resolver as diferenças de unidades dos dados, como: formatos, conversão de unidade, caracteres, entre outros (HAN; KAMBER, 2001).

b) *Discretização*: é definida como uma estratégia de redução de dados (DOUGHERTY et al., 1995), tem recebido crescente atenção e tornou-se uma etapa de pré-processamento utilizada em mineração de dados (GARCIA et al., 2013). O processo de discretização transforma atributos contínuos em atributos discretos. Isso é realizado associando-se cada intervalo de valores contínuos com um valor discreto (GARCIA et al., 2013).

### **3.1.3 Mineração de dados (MD)**

A mineração de dados trata-se de extrair ou minerar conhecimentos em grandes bases de dados. Esta é uma área cujo foco principal é desenvolver ou adaptar métodos e algoritmos para explorar um conjunto de dados produzidos e coletados em ambientes educacionais (BAKER; ISOTANI; CARVALHO, 2011).

A mineração de dados educacionais (MDE) busca compreender os dados

produzidos por alunos e professores, bem como suas interações por meio de diferentes ferramentas dentro de um AVEA.

A MDE pode ser aplicada, por exemplo, para descobrir os motivos que influenciam o aprendizado do aluno, analisar o comportamento do aluno evadido, entre outras possibilidades (BAKER; ISOTANI; CARVALHO, 2011; COSTA et al., 2012; WEBBER; ZAR; LIMA, 2013; RIGO et al., 2014).

Apesar da possibilidade de extração de relatórios, como dito na seção anterior, de forma a ajudar na tomada de decisão dos gestores e tutores, o número de dados é muito grande, o que torna quase impossível a análise desses relatórios sem a ajuda da tecnologia, nesse caso, da mineração de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

A MD compreende a aplicação de técnicas e algoritmos sobre os dados em busca de conhecimento útil e implícito (GOLDSCHMIDT; BEZERRA; PASSOS, 2015). Segundo Romero e Ventura (2013), a maioria das técnicas tradicionais de mineração de dados, tais como, classificação, agrupamento e técnicas de análise de associação, já foram aplicadas com êxito no domínio da educação.

A seguir, serão descritas, de forma sucinta, as técnicas de MD existentes na descoberta de conhecimento.

1) Associação: a associação apresenta padrões de relacionamento entre itens de uma base de dados. Ela busca os fatos que tendem a ocorrer de forma simultânea e com frequência. Poderíamos minerar os dados com base nas notas dos alunos em suas disciplinas, por exemplo, “90% dos alunos que têm bom desempenho na disciplina de Lógica têm bom desempenho em programação”. São exemplos de algoritmos que implementam essa técnica: Apriori, GSP, DHP, entre outros (ZAKI, 2000).

2) Classificação: classificação é o processo de encontrar um conjunto de modelos (funções) que descrevem e distinguem classes ou conceitos, com o propósito de utilizar o modelo para prever a classe de objetos que ainda não foram classificados, ou seja, ela busca prever uma classe de um novo dado automaticamente. Essa técnica é conhecida como predição, podendo ser de dois tipos: classificação ou estimação. Como exemplo, podemos utilizar um modelo de classificador para identificar quais são as principais causas da desistência de uma determinada disciplina. Com isso, podemos fazer conclusões, por exemplo, que a desistência do aluno está relacionada com a faixa etária entre 25 e 30 anos. Redes

Neurais, Árvore de Decisão, Algoritmos Genéticos e Lógica Nebulosa são exemplos de algoritmos que podem ser aplicados na classificação (CASTRO; FERRARI, 2016).

3) Regressão: a regressão é similar à classificação, porém é usada para valores numéricos, e não em uma categoria. A regressão, de acordo com Goldschmidt, Bezerra e Passos (2015), compreende a busca por uma função linear que mapeie os registros de uma base de dados em um intervalo de valores reais. Podemos, assim, estimar o valor de uma determinada variável analisando-se os valores dos demais registros. Enquanto isso, na tarefa anterior, os registros são classificados em uma classe, nessa tarefa, os registros são classificados em um valor baseado em uma função matemática (GARCIA, 2012). Entre as ferramentas que implementam a tarefa de regressão, estão: Estatística e Redes Neurais.

4) Agrupamento: como o próprio nome já diz, é a técnica de identificar e aproximar registros similares, também conhecidos como clusterização. É a coleção de registros similares entre si. Classificar, estimar ou prever não é a pretensão do agrupamento, ele serve para identificar os grupos de dados similares. Com essa técnica, espera-se conhecer novos atributos alvos (rótulos) a partir de um conjunto de dados, sem ter classificação prévia (COSTA, et al., 2013). Como exemplo, podemos buscar a construção do modelo de análise e tratamento da evasão e retenção discente em um determinado curso, permitindo reconhecer tipos de classes distintas de alunos por meio de dados similares dos alunos. Para a realização dessa técnica, podem ser utilizados os algoritmos: *k-Means*, *k-Modes*, *k-Prototypes*, Mapas de *Kohonen*, entre outros (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

5) Sumarização: é a técnica que permite a identificação de uma descrição compacta para os dados ou para um subconjunto deles. A sumarização visa identificar e indicar as características comuns entre um conjunto de dados. Essa tarefa é aplicada nos clusters obtidos na tarefa de clusterização ou segmentação. As técnicas de sumarização são, na maior parte das vezes, aplicadas à análise exploratória de dados e à geração automática de relatórios (GALVÃO, 2009). Alguns exemplos de algoritmos que implementam a técnica de sumarização são: Lógica Indutiva e Algoritmos Genéticos.

### 3.1.4 Avaliação de algoritmos

Após realizadas as etapas descritas anteriormente, devemos avaliar os resultados dos algoritmos. Neste trabalho, foi utilizada a técnica de avaliação cruzada, do inglês *cross-validation*, também conhecida como validação cruzada em *k*-pastas, do inglês *k-fold cross-validation*. Ela consiste na divisão da base de dados em *k* subconjuntos, sendo *k*-1 conjuntos para treinamento dos algoritmos e o conjunto restante para teste. O processo consiste em dividir o conjunto de dados de forma aleatória, para a validação do algoritmo em *N* subconjuntos de dados igualmente distribuídos em teste e treinamento. Esse processo é realizado para todos os *k* subconjuntos, e a média do desempenho para os conjuntos analisados é utilizada como indicador de qualidade de desempenho (CASTRO; FERRARI, 2016). A validação cruzada adotando *k*=10 é a mais utilizada para essa finalidade, sendo também utilizada neste trabalho.

Para a comparação de desempenho da predição, são empregadas as métricas de Acurácia, Precisão, Revocação e Medida-F. Para facilitar a compreensão dessas métricas, suas fórmulas serão demonstradas a seguir:

a) Acurácia (*accuracy*): porcentagem de amostras positivas e negativas classificadas corretamente sobre a soma de amostras positivas e negativas, ou seja, é a taxa de sucesso do classificador. A acurácia é formalmente definida como a porcentagem dos exemplos de teste que são corretamente classificados (SHALEVSHWARTZ; BEN-DAVID, 2014). Diante de um determinado conjunto de teste, é obtido o percentual dos exemplos que foram corretamente classificados. O cálculo de acurácia é dado por:

$$\text{Acurácia} = \frac{\text{Número de Classificações Corretas}}{\text{Número de Instâncias}}$$

Em que o Número de Classificações Corretas é a quantidade total de classificações previstas corretamente, tanto relacionadas aos alunos que foram classificados como desistentes quanto aos permanentes. Já o Número de Instâncias representa o número de alunos avaliados.

b) Precisão (*precision*): determina o percentual de acertos do algoritmo em relação à previsão realizada. O cálculo é dado por:

$$\text{Precisão} = \frac{\text{Nro.Alunos Previstos} \cap \text{Nro.Alunos Evadidos}}{\text{Nro.Alunos Previstos}}$$

Em que o Número de Alunos Previstos representa a quantidade de alunos que foi classificada, no âmbito da pesquisa, como “Evadido”, já que o objetivo é encontrar os alunos com tendência à evasão. Já o Número de Alunos Evadidos compreende a quantidade de alunos que, de fato, desistiu.

c) **Revocação (*recall*)**: mede a integridade do algoritmo em relação aos alunos evadidos, ou seja, descreve a porção que foi classificada corretamente como exemplos positivos. A medida é dada pela seguinte equação:

$$\text{Revocação} = \frac{\text{Nro.Alunos Previstos} \cap \text{Nro.Alunos Evadidos}}{\text{Nro.Alunos Evadidos}}$$

Em que o Número de Alunos Previstos representa a quantidade de alunos que foi classificada, no âmbito da pesquisa, como “Evadido”. Já o Número de Alunos Evadidos compreende a quantidade de alunos que, de fato, desistiu. Em suma, a métrica de revocação é calculada pelo número de alunos previstos e evadidos pelo número de alunos que evadiram.

d) **Medida-F (*f-Measure*)**: métrica de avaliação que consiste na média harmônica de Precisão e Revocação e é definida pela equação:

$$\text{Medida-F} = \frac{(2 * \text{Precisão} * \text{Revocação})}{(\text{Precisão} + \text{revocação})}$$

### 3.2 FERRAMENTA PARA MINERAÇÃO DE DADOS – WEKA

Entre as diversas ferramentas utilizadas para a mineração de dados, neste trabalho, optou-se pela utilização da ferramenta WEKA<sup>1</sup> (*Waikato Environment for Knowledge Analysis*), atualmente disponível na versão 3.6.3, mantida por universitários da Universidade de Waikato, na Nova Zelândia.

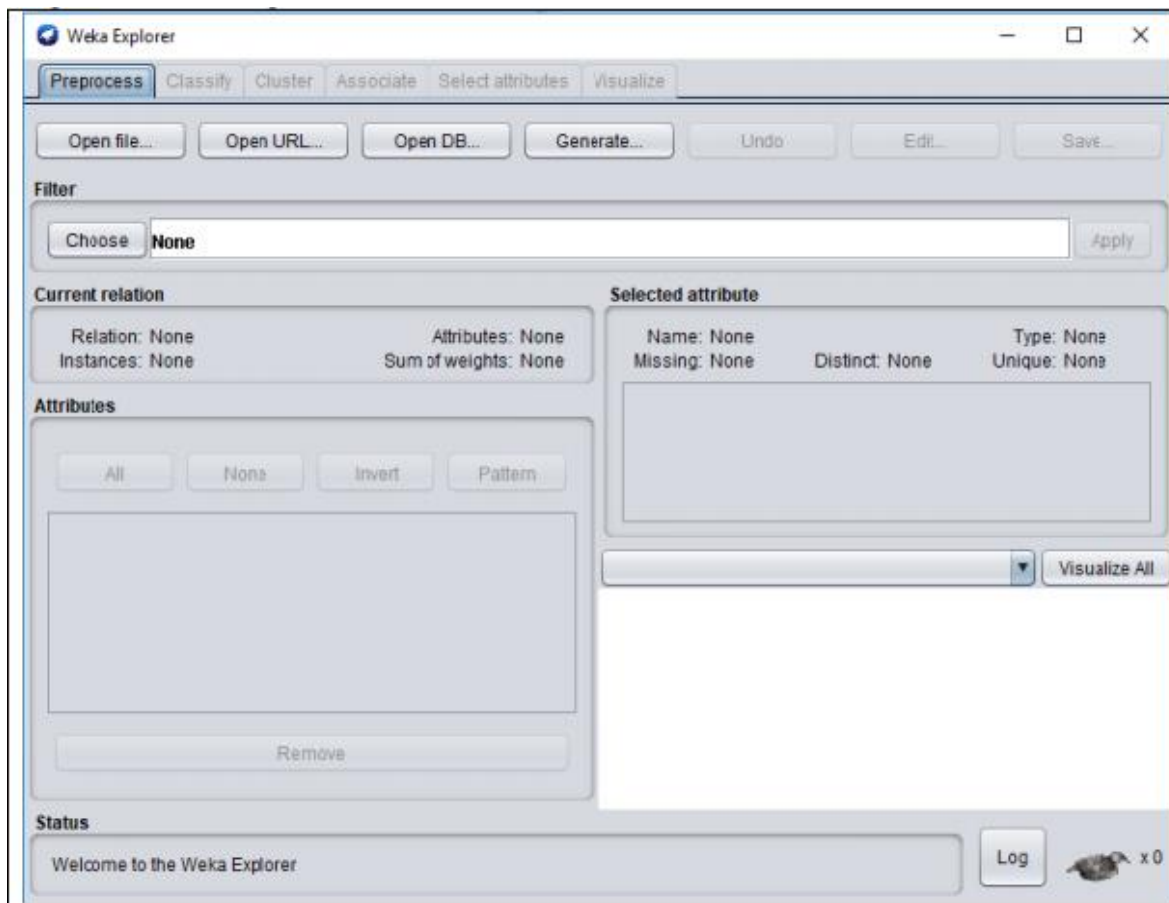
A WEKA é uma ferramenta desenvolvida utilizando a linguagem de programação JAVA e de código aberto, sob a *General Public License* (GNU). Pode ser utilizada tanto por meio de sua interface gráfica quanto em uma aplicação customizada através da importação de sua *Application Programming Interface* (API)

<sup>1</sup> <https://www.cs.waikato.ac.nz/ml/weka/>



(WEKA, 2019). Na figura 3, é possível visualizar a interface gráfica da ferramenta WEKA.

Figura 3 – Interface gráfica do WEKA



Fonte: Weka (2019).

As principais características da ferramenta WEKA são: o pré-processamento de dados, a visualização e análise preditiva, além de técnicas de modelagem, *listening*, agrupamento, associação, regressão e classificação, como já abordados anteriormente na seção anterior. No próximo capítulo, serão apresentados os trabalhos correlatos a esta pesquisa.

## 4 TRABALHOS CORRELATOS

Este capítulo busca apresentar brevemente algumas pesquisas relacionadas à evasão, bem como a análise dos trabalhos mencionados. A busca foi realizada por meio do Google Acadêmico, utilizando-se termos como: “mineração de dados educacionais”; “previsão de desempenho acadêmico”; “AVEA”; “AVA”; e “evasão EaD”.

### 4.1 DESCRIÇÃO DOS TRABALHOS

Rabelo *et al.* 2017, tomando como base as interações armazenadas no repositório do ambiente de pesquisa Moodle, encontraram 64 ações associadas a 24 módulos (atributos *action* e *module*), realizadas por 514 usuários com perfil de aluno em 13 Turmas de Cursos Graduação da UFRN, perfazendo um montante de 11.310.848 interações armazenadas no *log* do Moodle. Oito indicadores de desempenho foram selecionados: Ação de *login* de usuário (*login*); Ação de visualização do Curso (*view*); Ação de visualização de Recursos do Curso (*view*); Ação de Visualização de discussão em fórum (*view discussion*); Ação de Adição de postagem em fórum (*add post*); Ação de visualização de tarefa – *View*; Ação de enviar tarefa (*Submit*) e Ação de responder questionário (*Attempt*). Para aplicar as Técnicas de Mineração de Dados, fez-se o emprego da ferramenta computacional Weka. A técnica de classificação empregada nesse trabalho é a de “árvore de decisão”. Utilizaram-se dois algoritmos de classificação baseados em árvores de decisão, ID3 e *J48*. O melhor caso foi obtido pelo algoritmo *J48*, com 96,5% de acertos e 3,5% de erros de classificação. O algoritmo *J48* obteve acurácia de 96,5%, classificando corretamente o desempenho de 496 dos 514 alunos.

Burgos *et al.* (2017) propõem a utilização de mineração de dados e a criação de modelos de predição utilizando regressão logística linear para prever o risco de evasão de alunos. Foram utilizados dados de 104 alunos de diversos cursos na modalidade a distância. Os resultados apresentam valores de até 100% de acurácia geral já na quarta semana do curso. Segundo os autores, a aplicação dessa técnica junto a um plano de tutorial diminuiu em 14% o abandono escolar nos cursos em que foram aplicados.

Queiroga, Cechinel e Araújo (2017) apresentam uma proposta de metodologia para geração de modelos de predição de alunos em risco de evasão que utiliza a contagem de interações. Para o desenvolvimento da proposta, foram utilizados dados de quatro cursos técnicos na modalidade a distância do Campus Visconde de Graça (CaVG), do Instituto Federal Sul-rio-grandense (IFSul). Os autores utilizaram a contagem de interações dos estudantes no AVA como a principal informação para a geração dos modelos de predição. Para a avaliação dos modelos, foram utilizados cinco diferentes algoritmos: *Bayes Net*, *Simple Logistic*, *Multilayer Perceptron*, *Random Forest* e *J48*. A biblioteca utilizada foi o WEKA. Desde a primeira semana, as taxas de acerto foram superiores a 75% em quase todos os algoritmos, excetuando-se os resultados obtidos pelo *Simple Logistic*. Com o passar das semanas, as taxas de acerto elevaram-se para 87%. No segundo semestre do curso, foram obtidos resultados próximos a 94% antes do seu final, sendo possível afirmar, com quase 95% de exatidão, se um aluno irá terminar o curso antes do final do primeiro ano.

Detoni, Araujo e Cechinel (2015) buscaram detectar os alunos por meio da utilização da contagem de interações no AVEA Moodle em disciplinas na EaD. O experimento foi realizado na Universidade Federal de Pelotas (UFPEL), nos cursos de Licenciatura em Educação do Campo e Licenciatura em Pedagogia. A partir das interações, foram calculadas as médias de utilização do AVEA por parte dos alunos, tutores e professores e, posteriormente, foram aplicados os algoritmos de Redes Bayesianas, Redes Neurais, *J48* e *RandomForest*, em que foram obtidos resultados de até 67% de acurácia na predição do desempenho do aluno.

Nesse mesmo sentido, Rodrigues, Medeiros e Gomes (2013) tiveram como objetivo a obtenção de um modelo de previsão de desempenho de estudantes a partir de dados de interação durante o período de realização de um curso na modalidade a distância. O trabalho utilizou, para análise dos dados, o *software* estatístico *R*. Os autores afirmam, com uma confiança de 95%, que, para 79,2% dos alunos que interagem via fórum de discussão, à medida que aumenta a quantidade de interação, tende a aumentar o desempenho acadêmico.

Já Da Costa, Cazella e Rigo (2014) buscaram identificar o perfil de alunos com potencial de evasão. Para isso, utilizaram as notas e o número de interações no ambiente de cursos de especialização a distância de cada aluno. Para uso da ferramenta WEKA, esses dados foram transformados e posteriormente

classificados utilizando o algoritmo *J48* que obteve 97,6% de precisão.

Santana, Maciel e Rodrigues (2014) tiveram como objetivo realizar a avaliação da dimensão perfil de uso no ambiente Moodle. Os dados utilizados foram de 98 alunos de um curso semipresencial, extraídos do banco de dados do AVA Moodle. Os atributos utilizados foram: desempenho final (nota) e número de interações. Para analisar o desempenho do perfil dos alunos, foram utilizados 7 algoritmos, em que o *J48* obteve o melhor desempenho, alcançando 74% de acurácia.

Schmitt (2018) buscou apresentar aos gestores dados estratégicos por meio da mineração de dados educacionais, de modo que pudessem avaliar e gerar ações para mitigar a evasão na Universidade Federal de Santa Maria (UFSM). Sua pesquisa abrangeu dois experimentos, em três cursos de graduação, em que foram empregados dados de interações dos alunos no AVEA e dados do sistema de gestão acadêmico da UFSM. Para o primeiro experimento, foram utilizados exclusivamente os dados provenientes de logs do AVEA Moodle, sendo suas simulações efetuadas com dois formatos de dados, não normalizados e normalizados. Foram aplicados sete algoritmos diferentes, abrangendo dados sobre as interações dos alunos do curso de Letras – Espanhol. Todas as simulações realizadas com dados não normalizados apresentaram bons resultados, podendo-se observar que o melhor deles foi obtido pelo algoritmo *Naive Bayes*, com 58,39% de Medida-F. No entanto, algoritmos como o próprio *Naive Bayes*, *Random Forest* e *IBk* obtiveram, respectivamente, 51,59%, 54,34% e 43,56% de Medida-F para o formato normalizado.

## 4.2 ANÁLISE DOS TRABALHOS CORRELATOS

Ao analisar os resultados, observa-se uma maior utilização de algoritmos de Árvore de Decisão, fato comprovado pela literatura, por ser uma das formas mais simples e, ainda assim, mais bem-sucedidas de Algoritmos de Aprendizagem.

Outro ponto observado, na pesquisa, refere-se às ferramentas computacionais mais utilizadas. Nesse processo de descoberta de conhecimento, o Weka foi a de maior utilização nos trabalhos lidos e pesquisados.

Os objetivos das pesquisas são bem variados. As investigações abrangem estudos de desempenho em fóruns, em disciplinas específicas, em cursos, na

graduação a distância e presencial. Outras visam gerar alertas a estudantes com baixo desempenho ou com tendências à evasão, e têm-se, ainda, as que procuram traçar o perfil dos estudantes.

Entre as pesquisas, percebe-se a constante busca pelos motivos que levam os alunos a evadirem nas diferentes modalidades de ensino, no entanto é visível a falta de pesquisa nos cursos técnicos, foco desta pesquisa.

A utilização em larga escala do ambiente Moodle ficou evidente, sendo a maior fonte de dados para as pesquisas.

No que tange à classificação dos algoritmos com melhor taxa de acerto, o *J48* é um dos mais citados pelas pesquisas, obtendo altas taxas de acertos na maioria delas.

Para melhor sintetizar as diferenças entre os trabalhos analisados, elaboramos o quadro 3.

Portanto, após análise dos trabalhos, serão abordados, no próximo capítulo, os aspectos metodológicos utilizados na pesquisa, como forma de alcançarmos o objetivo proposto.

Quadro 3 – Análise dos trabalhos correlatos

<b>Autor</b>	<b>Modalidade</b>	<b>Técnica/Algoritmos utilizados</b>	<b>Ferramenta de KDD</b>	<b>Técnica de mineração utilizada</b>	<b>Atributos utilizados</b>
Rabelo et. al (2017)	Curso de graduação – EaD	ID3 e <i>J48</i>	WEKA	Classificação	<i>Login</i> ; visualização do curso, recursos, discussão em fórum, tarefas; adição de postagem; questionário
Burgos et. al (2017)	Cursos EaD – Não especificada a modalidade	<i>FEED Forward Neural Network (FFNN)</i> , <i>Support Vector Machine</i> , ARTMAP	Não especificada	Regressão	Contagem de interações no AVEA.
Queiroga, Araújo, Cechinel (2017)	Cursos técnicos EaD	Bayes Net, Simple Logistic, Multilayer Perceptron, Random Forest e <i>J48</i>	WEKA	Classificação	Contagem de interações no AVEA
Detoni, Araújo e Cechinel (2015)	Curso de graduação EaD	Redes <i>Bayesianas</i> , Redes Neurais, <i>J48</i> e <i>Random Forest</i>	WEKA	Classificação	Contagem de interações nas disciplinas no AVEA.
Rodrigues, Medeiros e Gomes (2013)	Curso de graduação EaD	Não especificado	estatístico R2	Regressão	Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem.
Da Costa, Cazella e Rigo (2014)	Curso de especialização EaD	<i>J48</i>	WEKA	Classificação	Nota; contagem de acesso ao ambiente
Santana, Maciel e Rodrigues (2014)	Curso semipresencial	<i>Random Forest</i> , <i>Multilayer Perceptron (MLP)</i> , <i>Naive Bayes</i> , SVM, KNN, <i>J48</i> e RBF	WEKA	Classificação	Nota; número de acesso ao fórum; interações e tempo médio de acesso ao ambiente
Schmitt (2018)	Curso de graduação EaD	<i>J48</i> , <i>Naive Bayes</i> , <i>MultiLayer Perceptron</i> , <i>Random Forest</i> , OneR, SMO e IBk.	WEKA	Classificação Precisão, Revocação e Medida-F	Contagem de interações. (Fórum, tarefas, materiais, links, pasta, <i>wiki</i> , questionário.)

Fonte: Autor (2020).

## 5 ASPECTOS METODOLÓGICOS

Inicialmente, foi adotado o procedimento de uma pesquisa exploratória acerca do tema norteador, buscando-se documentos, regulamentos do Instituto, dados e ações administrativas, taxas de evasão e experiências disponíveis na literatura.

Para o desenvolvimento da pesquisa, foi utilizado o Ambiente Virtual de Ensino-Aprendizagem Moodle, do Instituto Federal Farroupilha.

Após, por meio de uma pesquisa bibliográfica, buscaram-se os motivos da evasão nos cursos EaD e dados de evasão no Brasil, a evasão nos cursos EaD do IFFar e os conceitos dos AVEA. Para melhor entendimento sobre mineração de dados, foram abordados conceitos de KDD e MDE. A pesquisa tem uma abordagem, quali-quantitativa.

É quantitativa por utilizar dados provenientes da MD para deduzir características dos alunos que tendem a evadir, e qualitativa devido à compreensão em detalhes das causas relacionadas à evasão de alunos da modalidade EaD e da aplicação de técnicas de MD para identificação da evasão. Para a análise qualitativa, buscou-se apoio pedagógico da Diretoria de Educação a Distância.

Como forma de representar a relevância deste trabalho, buscaram-se trabalhos correlatos que propuseram a utilização de mineração de dados para a previsão da evasão de alunos na EaD.

Este trabalho foi realizado com base nas etapas do processo de KDD, descrita no capítulo 3 deste trabalho. Para a etapa de mineração de dados, foi utilizada a ferramenta WEKA descrita na seção 3.2. A escolha dessa ferramenta justifica-se pelos diversos estudos e abordagens na literatura, além das suas funcionalidades e características, tais como a facilidade de utilização, o alto número de algoritmos disponíveis para a utilização e a possibilidade de alteração dos parâmetros de execução destes.

Para esta pesquisa, foram utilizados dados acadêmicos dos cursos subsequentes de Administração e Agroindústria, ofertados na modalidade EaD pelo Programa Governamental da Rede e-Tec Brasil nos *campi* Santa Rosa e Alegrete do IFFar, respectivamente. A escolha desses cursos do ano de 2015 fundamenta-se devido a eles já terem sido finalizados; também pelo fato de os dados do SISTEC

dos cursos ofertados em 2017 não apresentarem evasão, não sendo possível utilizá-los para a mineração.

Os cursos analisados são ofertados em diversos polos de apoio presencial, possuindo disciplinas em sua estrutura curricular distribuídas em três semestres letivos, conforme projeto pedagógico de cada curso.

O curso de Administração utilizado foi ofertado pelo *campus* Santa Rosa em 10 polos e 4 *campi* e teve seu início no ano de 2015. Sua estrutura curricular é composta por 17 disciplinas em três semestres letivos, com carga horária total de 1.206 horas.

Já o curso de Agroindústria foi ofertado pelo *campus* Alegrete em 9 polos e teve seu início no ano de 2015. Sua estrutura curricular é composta por 16 disciplinas em três semestres letivos, com carga horária total de 1.005 horas.

Para o experimento 3, os dados das duas turmas foram agrupados em um único *dataset* para realizarmos a MD, formando um único conjunto de dados

Na figura 4, é mostrada uma representação gráfica com as etapas a serem seguidas na pesquisa.

Figura 4 – Etapas de desenvolvimento da pesquisa





## 6 DESENVOLVIMENTO DA PESQUISA

A realização desta pesquisa foi dividida em três experimentos: o primeiro, do curso de Administração; o segundo, do curso de Agroindústria; e, por último, a união das duas turmas.

### 6.1 PRIMEIRO EXPERIMENTO

Nesta seção, serão abordadas as etapas do processo de KDD, descritas na seção 3.1.

Para o primeiro experimento, foi utilizada a turma subsequente de Administração, ofertada no ano de 2015, conforme descrito na metodologia de pesquisa deste trabalho. É importante frisar a taxa de evasão do curso em questão, conforme dados do SISTEC, de 42,2%.

#### 6.1.1 Seleção dos dados

Os dados utilizados na pesquisa estão armazenados em duas fontes distintas. No AVEA Moodle, estão os dados das interações dos alunos no ambiente, os quais consistem de *LOG's* gerados pelo sistema durante cada ação realizada pelo estudante no ambiente.

Os dados foram disponibilizados com uma cópia do banco de dados do ambiente pela empresa que hospeda o serviço, devido a este ser terceirizado pela instituição, sendo que a autorização para a sua utilização foi obtida junto à Reitoria, conforme pode ser visto nos anexos A e B deste trabalho.

Essa etapa iniciou pela extração dos dados do Moodle via linguagem *SQL*, que consiste em consultas diretamente no banco de dados. Essas consultas precisam ser montadas com base nas tabelas do banco de dados e diferenciam-se pela versão do Moodle utilizado. No caso do IFFar, a versão utilizada é a 3.1.3. Vale informar que, para que, de fato, pudéssemos utilizar os dados para a mineração, várias consultas e análises de dados fizeram-se necessárias. Para a realização da extração dos dados, foram utilizados todos os dados disponíveis na tabela do banco de dados *mdl\_logstore*, no qual se encontram armazenadas todas as interações realizadas no AVEA. Os dados extraídos podem ser visualizados no quadro 4.

Quadro 4 – Dados brutos extraídos do AVEA

A	B	C	D	E
curso	disciplina	componente	acao	contagem
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_assign	viewed	2
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_assign	viewed	1
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	reviewed	1
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	reviewed	2
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	started	1
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	started	2
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	submitted	1
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	submitted	2
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	viewed	22
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_quiz	viewed	82
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_resource	viewed	32
Técnico em ADMINISTRAÇÃO	Administração Financeira	mod_resource	viewed	11
Técnico em ADMINISTRAÇÃO	Ambientação em EAD (E)	mod_assign	viewed	9
Técnico em ADMINISTRAÇÃO	Ambientação em EAD (E)	mod_assign	viewed	1
Técnico em ADMINISTRAÇÃO	Ambientação em EAD (E)	mod_assign	viewed	1
Técnico em ADMINISTRAÇÃO	Ambientação em EAD (E)	mod_forum	created	2
Técnico em ADMINISTRAÇÃO	Ambientação em EAD (E)	mod_forum	uploaded	1

Fonte: Autor (2020).

No banco de dados do AVEA MOODLE, encontram-se atributos que precisam ser analisados pela ferramenta WEKA, utilizada neste trabalho, para buscar os que melhor ajudam a caracterizar o aluno evadido; os atributos disponíveis após a extração dos dados podem ser vistos no quadro 5.

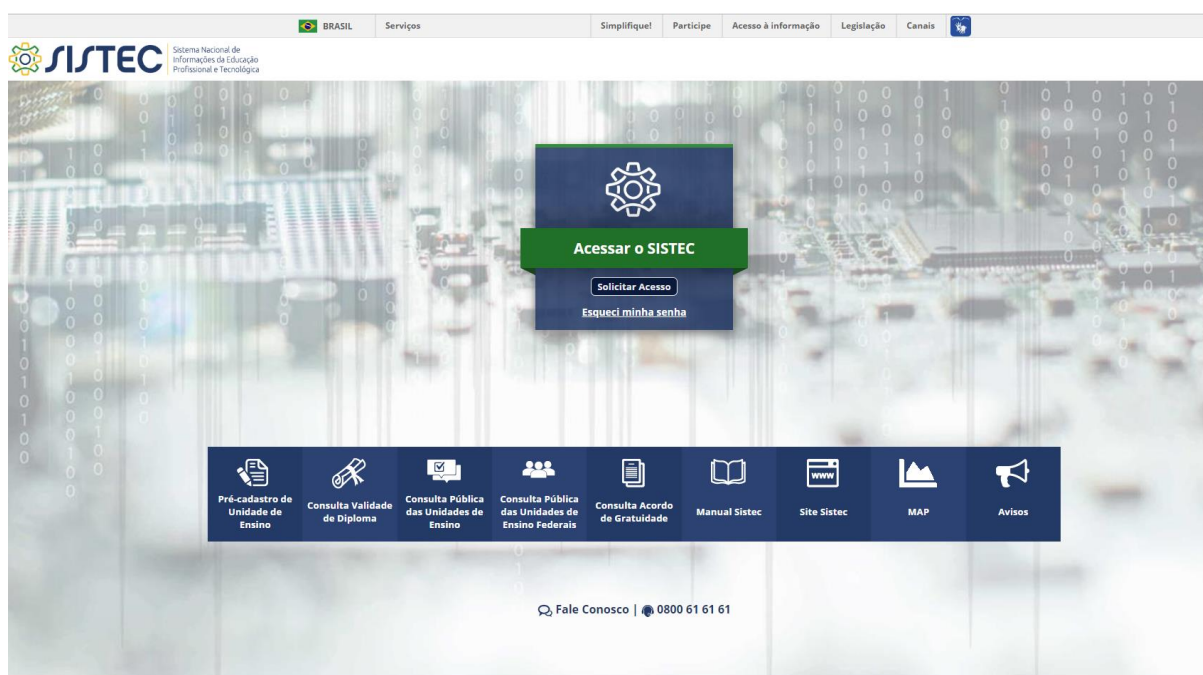
Quadro 5 – Atributos do AVEA Moodle

Recurso	Atributo	Descrição
Fórum	forum_add	Número de postagens em fóruns.
	forum_view	Número de visualizações nos fóruns.
Tarefas	assign_submit	Número de tarefas submetidas.
	assign_view	Número de visualizações de tarefas.
Materiais	resource_view	Número de visualizações de materiais disponibilizados.
Wiki	wiki_view	Número de interações no wiki.
Questionário	quiz_view	Número de interações nos questionários.
Links	url_view	Número de visualizações de links
Chat	Chat_view	Número de visualizações em chat
	Chat_sent	Número de mensagens enviadas
Pasta	Folder_viewed	Número de pastas visualizadas

Fonte: Autor (2020).

A outra fonte de dados (Figura 5) é o Sistema Nacional de Informações da Educação Profissional e Tecnológica (SISTEC), que contém as informações do aluno, bem como sua situação de matrícula escolar dentro da instituição, podendo ser concluído ou evadido.

Figura 5 – Interface web do SISTEC



Fonte: <<https://sistec.mec.gov.br/login/login>>.

Para obtermos dados dessa segunda fonte, foi necessário solicitar ao setor de Pesquisa Institucional sua extração. Os dados foram disponibilizados em formato de planilha, contendo dados pessoais do aluno, como por exemplo número de identidade, sexo, CPF, situação no curso matriculado, concluído ou em abandono, etc.

### 6.1.2 Preparação dos dados

Após os dados serem selecionados e extraídos, resultando em mais de 14 mil linhas, conforme pode ser visto no quadro 4, foi realizada a preparação ou, como é conhecido, pré-processamento de dados, com os componentes e as ações que continham alguma interação. Essa seleção de componentes e ações deu-se por meio da visualização da contagem do número de interações em cada componente

pelos alunos. Sendo assim, componentes e ações em que não existia interação alguma foram desconsiderados devido à não utilização do recurso durante os três semestres. A figura 6 mostra os dados preparados.

Figura 6 – Dados preparados

B	C	D	E	F	G	H	I	J
assign_submitted	assign_viewed	chat_sent	chat_viewed	folder_viewed	forum_viewed	resource_viewed	url_viewed	Classe
20	447	7	22	15	192	254	18	Concluído
23	256	1	32	11	218	572	13	Concluído
15	133	4	15	7	105	495	16	Concluído
7	188	0	0	12	131	292	12	Concluído
10	118	0	0	8	76	168	17	evadido
26	242	25	9	15	112	933	40	Concluído
13	73	0	2	1	42	102	0	Concluído
6	37	0	0	0	22	61	2	evadido
0	0	0	0	1	0	8	0	evadido
1	14	0	0	0	25	52	0	evadido
0	0	0	0	0	0	1	0	evadido
0	0	0	0	0	5	15	0	evadido

Fonte: Autor (2020).

Podemos observar que possuímos uma linha de dados por aluno, com a contagem total de interações considerando os atributos selecionados.

O próximo passo foi remover os dados dos alunos que não continham interação alguma durante todo o curso, ou seja, alunos sem qualquer tipo de interação dentro do ambiente foram removidos como forma de não prejudicar a predição pelos algoritmos. A justificativa para não haver interação, segundo a Diretoria de Educação a Distância, é devido a esses alunos nunca terem acessado o Moodle. Portanto, para este trabalho, eles não fazem parte do cálculo da evasão. Nesse curso, após a preparação dos dados, a planilha teve um total de 180 instâncias.

Entre os atributos selecionados, para que possamos caracterizar os dados, estes foram classificados em duas classes: “concluído” e “evadido”. A primeira classe continha 104 instâncias e representa os alunos que concluíram o curso; já na classe “evadido”, com 76 instâncias, continha alunos que, de fato, evadiram, conforme fonte de dados do SISTEC.

Para a realização da mineração nos três experimentos, após a qualificação deste trabalho, foi verificado que o atributo Quiz estava tendencioso, devido ao número alto de questionários abertos para as turmas. Não sendo possível ponderar esses dados, o atributo Quiz foi desconsiderado.

Os demais atributos não utilizados, não continham interação e por isso, não foram considerados para a mineração nos 3 experimentos.

Como forma de visualizar os dados finais para a mineração, no quadro 6, constam os atributos utilizados para a mineração.

Quadro 6 – Atributos e dados finais utilizados na MDE

Assign_submitted – Tarefa submetida
Assign_viewed – Tarefa visualizada
Chat_sent – Mensagem enviada
Chat_viewed – Mensagem visualizada
Folder_viewed – Pasta visualizada
Fórum_viewed – Visualização no fórum
Resource_viewed – Material visualizado
url_viewed – Link visualizado
Classe do aluno (evadido, concluído)

Fonte: Autor (2020).

Como próxima etapa, os dados foram transformados para o formato de Arquivo de Relação de Atributos, do inglês *Attribute-Relation File Format* (ARFF), conforme figura 7, para que fossem utilizados pela ferramenta WEKA.

Figura 7 – Arquivo ARFF

```

@relation treinamento
@attribute tarefa_submetida numeric
@attribute tarefa_visualizada numeric
@attribute chat_sent numeric
@attribute chat_viewed numeric
@attribute folder_viewed numeric
@attribute forum_viewed numeric
@attribute quiz_viewed numeric
@attribute quiz_started numeric
@attribute quiz_submitted numeric
@attribute quiz_reviewed numeric
@attribute resource_viewed numeric
@attribute url_viewed numeric
@attribute class {concluido, evadido}
@data

20,447,7,22,15,192,975,96,94,74,254,18,concluido
23,256,1,32,11,218,789,98,97,110,572,13,concluido
15,133,4,15,7,105,543,71,70,116,495,16,concluido
7,188,0,0,12,131,478,60,58,8,292,12,concluido
10,118,0,0,8,76,88,11,11,10,168,17,evadido

```

Fonte: Autor (2020).

### 6.1.3 Mineração de dados: primeiro experimento

Após a seleção dos atributos e preparação dos dados, ocorre a etapa de modelagem, por meio de testes e implementações dos algoritmos. A seleção dos algoritmos que foram utilizados deu-se por meio dos trabalhos correlatos e de sua análise, bem como de sua disponibilidade na ferramenta WEKA.

Dessa forma, foram escolhidos os seguintes algoritmos para a realização dessa etapa: *J48*, *Naive Bayes*, *MultiLayer Perceptron*, *Randon Forest*, *IBK*, *SMO*, *OneR*.

Para a geração dos modelos, foi utilizado o método *cross-validation*, assumindo um valor de 10 pastas. Na tabela 2, é possível visualizar os resultados dos testes realizados por meio das métricas de avaliação Acurácia, Precisão, Medida-F e Revocação.

Tabela 2 – Métricas de desempenho dos algoritmos utilizados: primeiro experimento

Algoritmo	Métricas de avaliação			
	Acurácia	Precisão	Medida-F	Revocação
<i>J48</i>	91,66%	91,70%	91,70%	91,70%
<i>Naive Bayes</i>	83,88%	84,50%	84,00%	83,90%
<i>MultiLayer Perceptron</i>	90,00%	90,10%	90,00%	90,00%
<i>Randon Forest</i>	93,33%	93,33%	93,33%	93,33%
<i>IBK</i>	90,00%	90,00%	90,00%	90,00%
<i>SMO</i>	92,22%	92,30%	92,20%	92,20%
<i>OneR</i>	92,77%	93,10%	92,70%	92,80%

Fonte: Autor (2020).

Ao analisarmos a tabela 2, podemos perceber que os dados são semelhantes entre os algoritmos, com uma alta taxa de acerto, sendo satisfatório para todos os algoritmos selecionados, apresentando uma média de acurácia superior a 90%.

O algoritmo *Naive Bayes* apresentou o menor índice de acerto para todas as métricas de avaliação, com taxa de 83,88% para a acurácia.

Por fim, o algoritmo com melhor taxa foi o *Randon Forest*, apresentando 93,33% em todas as métricas de avaliação.

A figura 8 mostra os dados da árvore de decisão, gerada pelo algoritmo *J48*, em porcentagem. Conforme o número de interações em cada recurso, os valores foram transformados para porcentagem, considerando o número máximo de

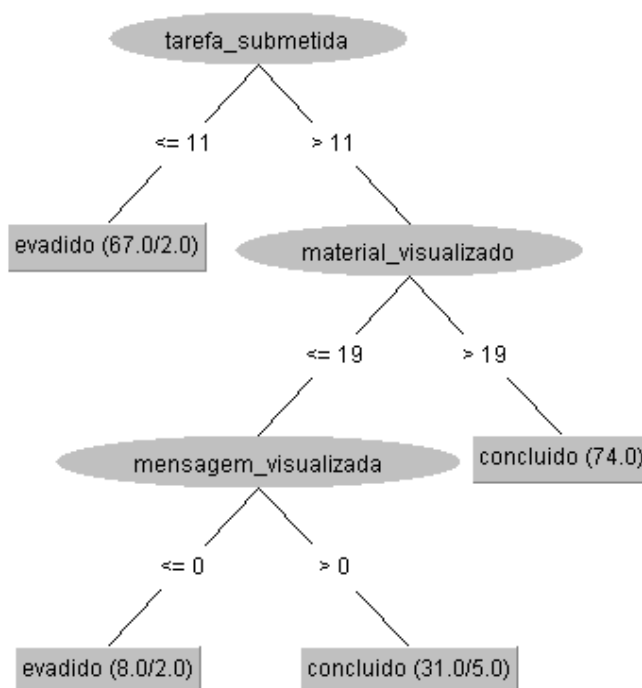
interações como 100%. A tabela 3 mostra o valor máximo de interação em cada recurso utilizado, ao longo dos 3 semestres do curso.

Tabela 3 – Interação máxima: primeiro experimento

Recurso	Interação máxima
Tarefa submetida	44
Tarefa visualizada	447
Chat enviado	41
Chat visualizado	80
Pasta visualizada	78
Fórum visualizado	462
Material visualizado	1485
URL visualizado	87

Fonte: Autor (2020).

Figura 8 – Árvore de decisão em porcentagem: primeiro experimento



Fonte: Autor (2020).

Conforme observado na árvore de decisão, entre os atributos analisados, o atributo que determinou a evasão do aluno foi o recurso tarefa submetida. Sendo assim, ela nos mostra que os 67 alunos que submeteram apenas 11% ou menos do número máximo de tarefas, o algoritmo classificou como evadido, e para os alunos que submeteram mais de 11% das tarefas, o algoritmo passou a considerar a visualização de material. Os 74 alunos que visualizaram mais de 19% do número máximo de materiais disponibilizados, concluíram o curso. Para quem visualizou 19% ou menos, o algoritmo analisou a visualização de chat. Oito alunos não visualizaram vez alguma e evadiram, e 31 alunos que visualizaram ao menos uma mensagem concluíram o curso.

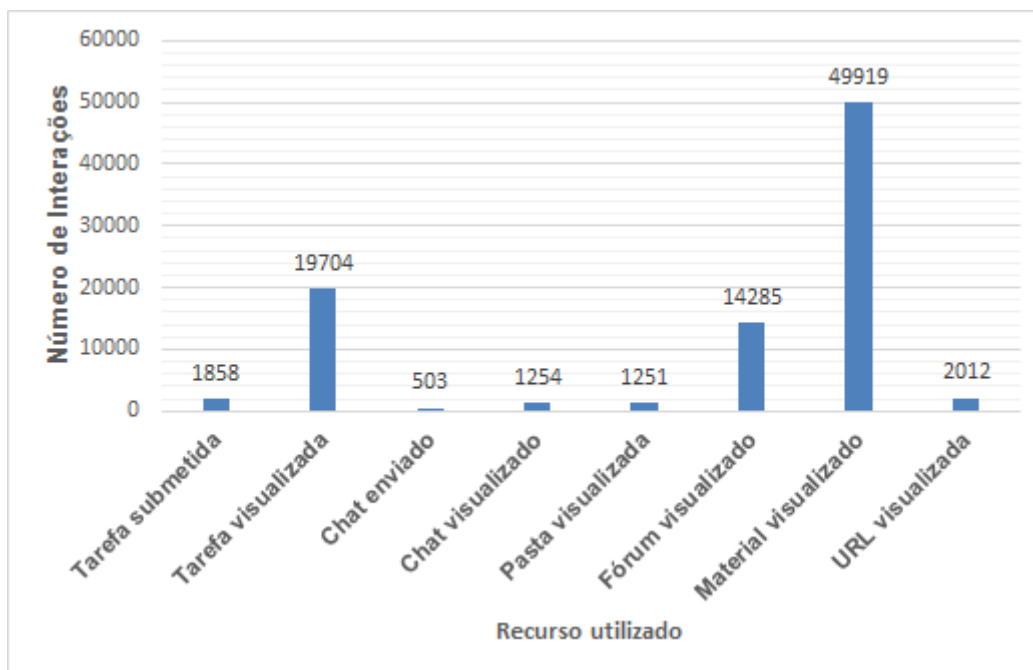
Na figura 8 devemos atentar para os números entre parênteses abaixo de cada classificação, eles nos mostram primeiramente quantos alunos foram classificados pelo algoritmo naquela classe e ao lado, quantos ele classificou incorretamente.

Dessa forma, os dois recursos mais utilizados para prever a evasão do aluno, para esse curso de Administração, foram tarefa submetida e visualização de material.

As interações em cada atributo, somadas durante todo o período do curso, independentemente da classe evadido e concluído, podem ser mais bem visualizadas no gráfico 3.



Gráfico 3 – Interações totais nos recursos utilizados do AVEA: primeiro experimento

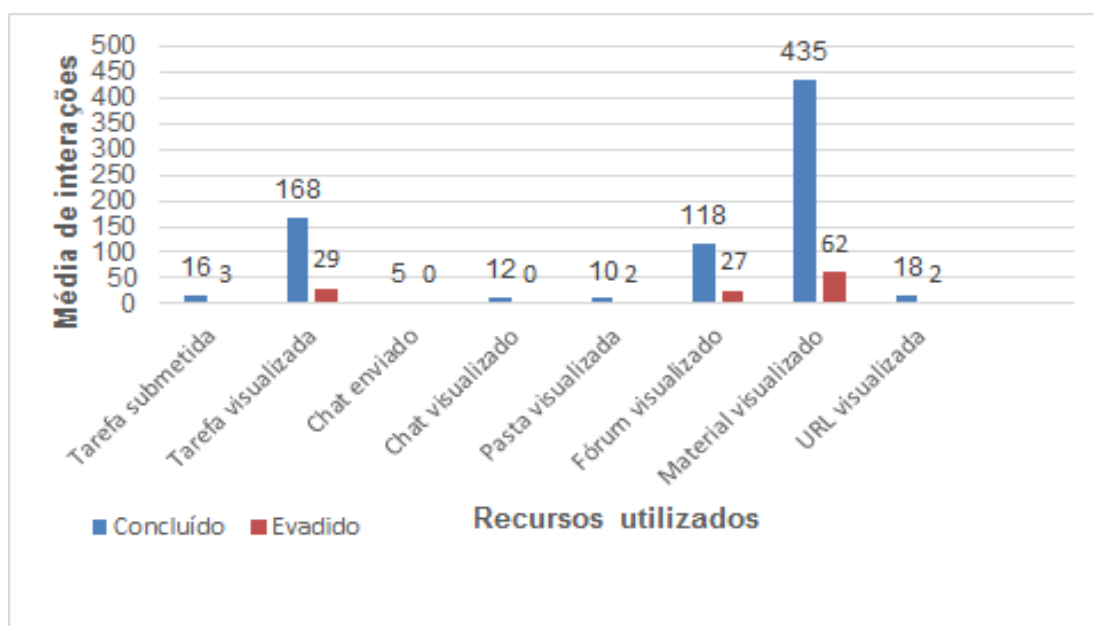


Fonte: Autor (2020).

Podemos observar que o recurso mais utilizado no curso de Administração, ofertado em 2015 pelo *campus* Santa Rosa, é o recurso material visualizado, seguido da visualização de tarefas. Os recursos com menor interação foram o *chat*, o acesso às pastas, a tarefa submetida e os *links* disponibilizados. Esses dados podem variar conforme a metodologia de aula de cada professor em diferentes turmas.

Por fim, no gráfico 4, é possível visualizar a média de utilização dos recursos do AVEA que foram empregados no processo de MD, considerando interações de 104 alunos concluintes e de 76 evadidos. Os dados estão distribuídos em duas classes: concluído e evadido. Para os dados gerados no gráfico 4, foram consideradas as interações de cada classe, durante todo o curso, em cada recurso utilizado, e divididas pelo número de alunos correspondente a cada classe.

Gráfico 4 – Média de utilização dos recursos no AVEA: Primeiro experimento



Fonte: Autor (2020).

Podemos analisar que o recurso tarefa submetida tem uma média de 16 interações para os alunos que concluíram o curso e 3 interações para os alunos que evadiram, evidenciando ainda mais a árvore de decisão gerada pelo algoritmo *J48*. O mesmo ocorre com os demais recursos, ficando evidente a diferença média na utilização dos recursos entre os alunos evadidos e concluídos.

#### 6.1.4 Análise dos dados: primeiro experimento

Ao analisar os dados qualitativamente, deduz-se que os estudantes mantêm maior frequência virtual na realização das tarefas avaliativas e no estudo do material, o que indica um padrão de uso bastante objetivo na conclusão das tarefas avaliativas, visto que outras interações possuem menor atenção. Por exemplo, o chat – que é utilizado para a mediação pedagógica entre estudante e conteúdo e requer participação síncrona – possui pouca adesão dos estudantes, que não estão acostumados a destinar um tempo maior para os estudos que não envolvem processos avaliativos.

Outros aspectos também reforçam essa cultura de utilização do AVEA Moodle. No curso de Administração, a ação nas disciplinas resume-se,

basicamente, em arquivos de texto e/ou vídeo e quiz, fato que pode ser visualizado nas árvores. É possível que, em decorrência disso, o quiz tenha sido tão expressivamente utilizado, a ponto de tornar-se tendencioso na amostra, pois existem disciplinas que o utilizaram como única tarefa disponibilizada durante todo o semestre, tanto nas tarefas ao longo do curso quanto nas avaliações finais.

Essa análise é mais bem compreendida por Almeida et al. (2013, p. 20), ao destacarem que a evasão depende de questões culturais: “ensinar e aprender a distância não são tarefas fáceis e ambos os atores diretamente implicados – professor e aluno – precisam passar por uma mudança cultural”.

## 6.2 SEGUNDO EXPERIMENTO

Para a realização desse experimento, foi utilizado o curso de Agroindústria, ofertado pelo *campus* Alegrete, em 2015. As etapas de seleção e preparação dos dados deram-se da mesma forma do primeiro experimento, descritas no item 6.1.1 e 6.1.2. Sendo assim, elas não serão descritas nos próximos dois experimentos. A taxa de evasão desse curso foi de 51,55%. No experimento, o número total de instâncias foi 161. A classe concluída continha 78 instâncias, já a classe “evadido”, 83 instâncias, conforme fonte de dados do SISTEC.

### 6.2.1 Mineração de dados: segundo experimento

De forma a podermos comparar os dados entre as turmas, mantivemos os mesmos algoritmos utilizados no primeiro experimento.

Tabela 4 – Métricas de desempenho dos algoritmos utilizados: segundo experimento

Algoritmo	Métricas de avaliação			
	Acurácia	Precisão	Medida-F	Revocação
<i>J48</i>	85,71%	88,44%	85,50%	85,70%
<i>Naive Bayes</i>	79,50%	81,50%	79,10%	79,50%
<i>MultiLayer Perceptron</i>	84,47%	84,50%	84,5%	84,50%
<i>Randon Forest</i>	80,74%	80,80%	80,70%	80,70%
<i>IBK</i>	83,22%	83,20%	83,20%	83,20%
<i>SMO</i>	81,98%	83,80%	81,70%	82,20%
<i>OneR</i>	83,22%	84,90%	83,10%	83,20%

Fonte: Autor (2020).

Da mesma forma que o primeiro experimento, nessa mineração, os dados apresentaram bons resultados. Para ambas as métricas de avaliação dos algoritmos, as taxas foram altas, tendo a maior taxa de acerto, a acurácia, obtida com o algoritmo *J48*, de 85,71%.

No que diz respeito à taxa de acertos dos algoritmos em relação à previsão realizada, a precisão, os melhores índices também foram desse algoritmo, com taxa de 88,44%.

A medida-F, que combina os resultados de precisão e revocação, não foi diferente, alcançando boa taxa de 85,50%.

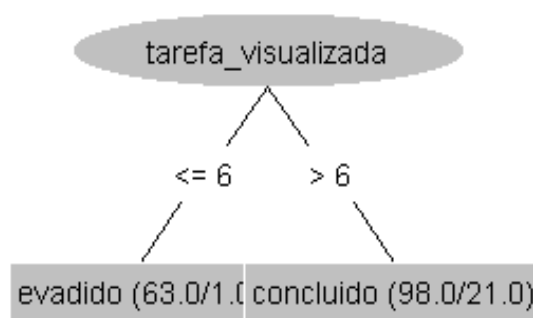
Na figura 9, visualizamos a árvore de decisão gerada em porcentagem pelo algoritmo *J48*. Para essa árvore, consideramos, na tabela 5, as seguintes interações máximas em cada recurso:

Tabela 5 – Interação máxima: segundo experimento

<b>Recurso</b>	<b>Interação máxima</b>
Tarefa submetida	36
Tarefa visualizada	544
Chat enviado	0
Chat visualizado	1
Pasta visualizada	22
Fórum visualizado	1639
Material visualizado	1822
URL visualizado	298

Fonte: Autor (2020).

Figura 9 – Árvore de decisão em porcentagem: segundo experimento



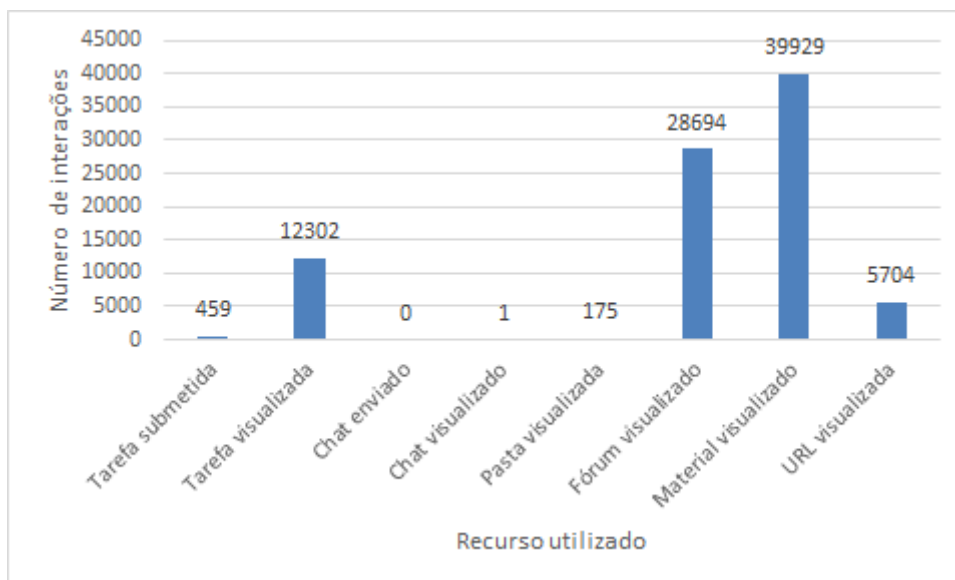
Fonte: Autor (2020).

Podemos observar que a árvore foi gerada com um único nó, e o recurso que determinou a evasão pelo algoritmo foi tarefa visualizada. Esse fato, de início, reflete a pouca exploração dos recursos e tarefas no AVEA Moodle.

Nesse cenário, entende-se que o aluno que interagiu 6%, ou menos, do total de tarefas disponibilizadas, evadiu, contabilizando 63 alunos evadidos no curso. Já aqueles que interagiram mais de 6% do número máximo de visualização de tarefas, concluíram o curso, sendo 98 alunos classificados nessa classe. Porém, como estamos falando de previsão, podemos perceber, também, que, dos 63 alunos que foram classificados pelo algoritmo como evadido, um foi classificado de forma incorreta, e, dos 98 alunos que o algoritmo classificou como concluído, 21 resultados apresentaram-se incorretos. Esses dados mostram a taxa de acertos do algoritmo, de 85,71%.

No gráfico 5, observa-se a utilização dos recursos utilizados durante todo o curso.

Gráfico 5 – Interações totais nos recursos utilizados do AVEA: segundo experimento

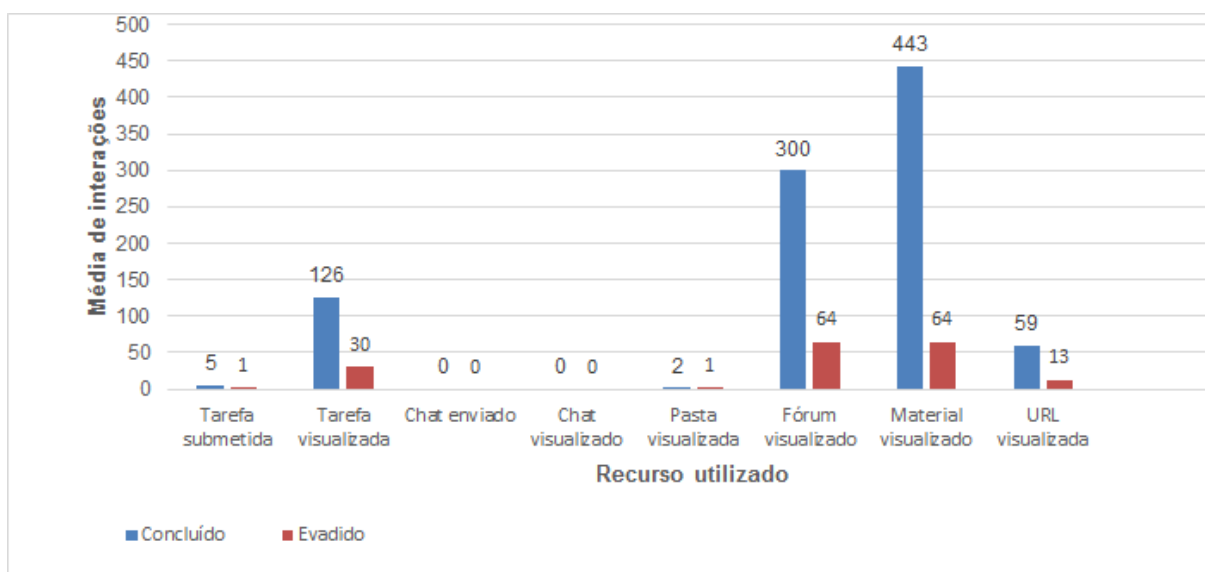


Fonte: Autor (2020).

Observa-se que o recurso mais utilizado no curso de Agroindústria, ofertado em 2015 pelo Campus Alegrete, é o recurso material visualizado, seguido da visualização no fórum, o que, somado ao quiz, caracterizava a estrutura das salas virtuais dos cursos da Rede e-Tec. Observa-se que o recurso chat praticamente não foi utilizado, assim como outros recursos e tarefas possíveis.

Por fim, no gráfico 6, é possível visualizar a média de utilização dos recursos do AVEA que foram empregados no processo de MD, considerando interações de 78 alunos concluintes e de 83 evadidos. A metodologia de apresentação dos dados foi a mesma utilizada no experimento 1.

Gráfico 6 – Média de utilização dos recursos no AVEA: segundo experimento



Fonte: Autor (2020).

Conforme os dados do gráfico 6, observam-se os dois recursos mais utilizados, material visualizado e fórum visualizado. A média de interação do recurso material para os alunos concluintes foi de 443 visualizações. Já para os alunos evadidos, a média de visualização é 64. Para o recurso fórum, a média de visualização pelos alunos concluintes é 300, para os alunos evadidos, também foi de 64.

### 6.2.2 Avaliação dos dados: segundo experimento

Nesse curso, percebe-se como baixa a utilização dos recursos e tarefas do Moodle como ambiente virtual de ensino-aprendizagem. O AVEA necessita ser explorado em sua complexidade e cabe ao professor buscar maneiras de realizar a acolhida e interação com o estudante, por meio da transposição didática dos conteúdos. Por isso, entende-se por que, no referido curso, a árvore foi gerada com apenas um recurso, colaborando para o entendimento da pouca fluência dos docentes e tutores no ambiente, no que compete à didática EaD. Observa-se, também, a não utilização do recurso chat, evidenciando ainda mais a pouca exploração de recursos e atividades síncronas.

Ainda, ao analisar o ambiente Moodle, percebe-se que a interação nos fóruns

é mínima, por parte dos tutores e professores, o que também não provoca o estudante a participar. Conforme aponta Mill (2014, p. 25), na EaD:

Cabe a diferentes profissionais as tarefas de produzir o conteúdo do curso, de organizar didaticamente o material, de converter o material para a linguagem da mídia (impressa, audiovisual, virtual etc.) de coordenar todas atividades de um curso e manejar/gerenciar a turma, entre outras.

Outro fato importante a ser destacado é que o atributo nota não seria possível de utilização, pois as avaliações, nesse curso, eram realizadas em papel, não no AVEA, enfatizando ainda mais o restrito fluxo de uso do AVEA, tanto pelos estudantes quanto pelos professores e tutores.

Entretanto, não queremos afirmar, com isso, que o trabalho não possuía seu grau de qualidade. No âmbito da Rede e-Tec, os materiais didáticos eram criados especificamente para os cursos, os professores e tutores foram selecionados por editais, e, ainda, o trabalho era acompanhado por uma equipe de apoio técnico, pedagógico e administrativo. O fator complicador, nesse cenário, talvez, seja a quantidade de ofertas ter crescido de modo mais expressivo que o tempo de a instituição preparar-se para tal, causando algumas fragilidades nos processos, por mais que houvesse grande esforço das equipes multidisciplinares, demanda de estudantes e potência social no Programa Governamental.

### 6.3 TERCEIRO EXPERIMENTO

Nesse cenário, os dados dos dois cursos foram agrupados, formando um único conjunto de dados, ou seja, um mesmo *dataset*. O conjunto teve 341 instâncias, e o objetivo maior foi poder unir os recursos e interações em uma única base, de forma a buscar dados genéricos.

#### 6.3.1 Mineração de dados: terceiro experimento

Foram consideradas as mesmas métricas de avaliação dos experimentos anteriores, os resultados podem ser vistos na tabela 6.



Tabela 6 – Métricas de desempenho dos algoritmos utilizados: terceiro experimento

Algoritmo	Métricas de avaliação			
	Acurácia	Precisão	Medida-F	Revocação
<i>J48</i>	82,69%	82,70%	82,70%	82,70%
<i>Naive Bayes</i>	84,45%	86,00%	84,40%	84,50%
<i>MultiLayer Perceptron</i>	85,63%	85,70%	85,60%	85,60%
<i>Randon Forest</i>	88,26%	88,30%	88,30%	88,30%
<i>IBK</i>	87,39%	87,40%	87,40%	87,40%
<i>SMO</i>	87,39%	87,70%	87,40%	87,40%
<i>OneR</i>	84,23%	81,20%	81,20%	81,20%

Fonte: Autor.

Nessa mineração, a maior taxa de acerto, a acurácia, foi do algoritmo *Random Forest*, de 88,26%. Os demais algoritmos também apresentaram boas taxas, com taxa média de acurácia em 85,72%.

A acurácia apresentou menor índice para o algoritmo *J48*, com 82,69%, porém, para as demais métricas de avaliação, a menor taxa foi do algoritmo *OneR*, com 81,20%.

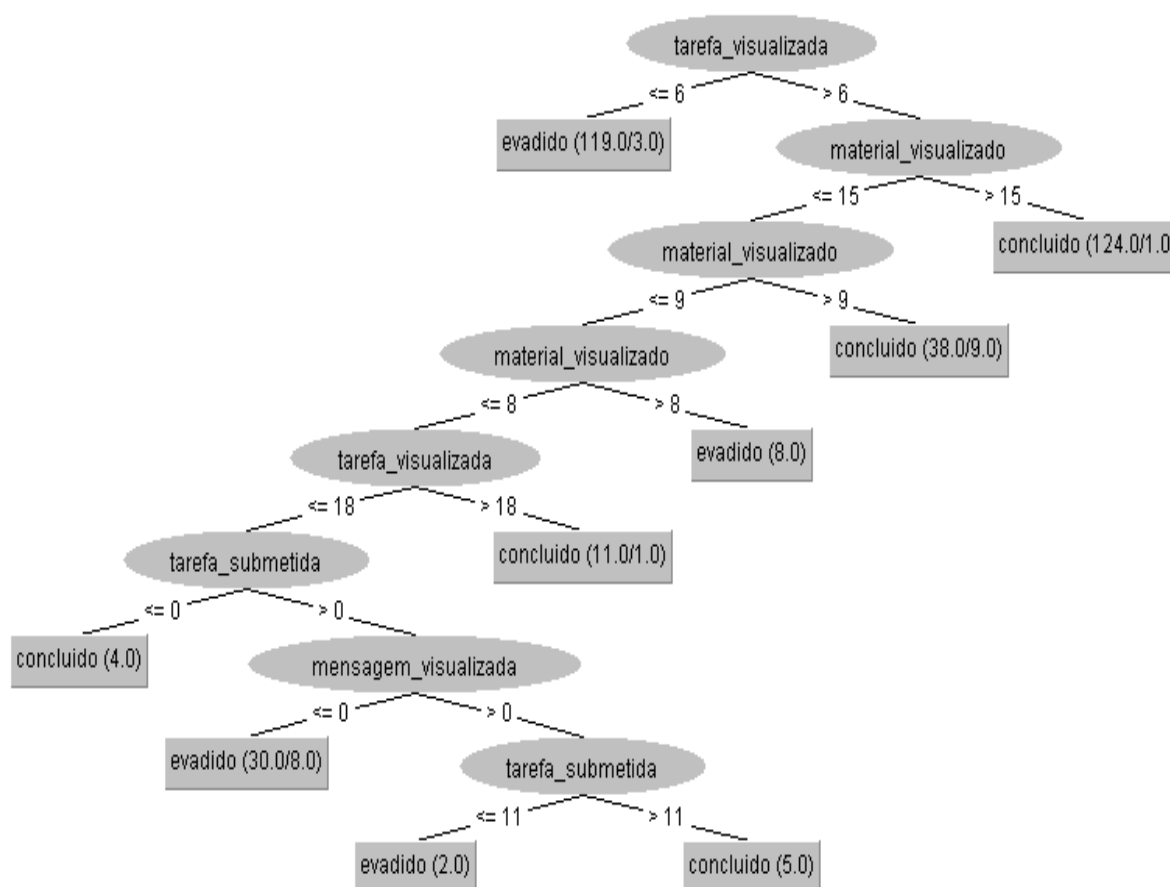
A seguir, a tabela 7 apresenta o valor de interação máxima utilizado em cada recurso, e a figura 10, a árvore de decisão gerada em porcentagem.

Tabela 7 – Interação máxima: terceiro experimento

Recurso	Interação máxima
Tarefa submetida	44
Tarefa visualizada	544
Chat enviado	41
Chat visualizado	80
Pasta visualizada	78
Fórum visualizado	1639
Material visualizado	1822
URL visualizado	298

Fonte: Autor (2020).

Figura 10 – Árvore de decisão em porcentagem: terceiro experimento



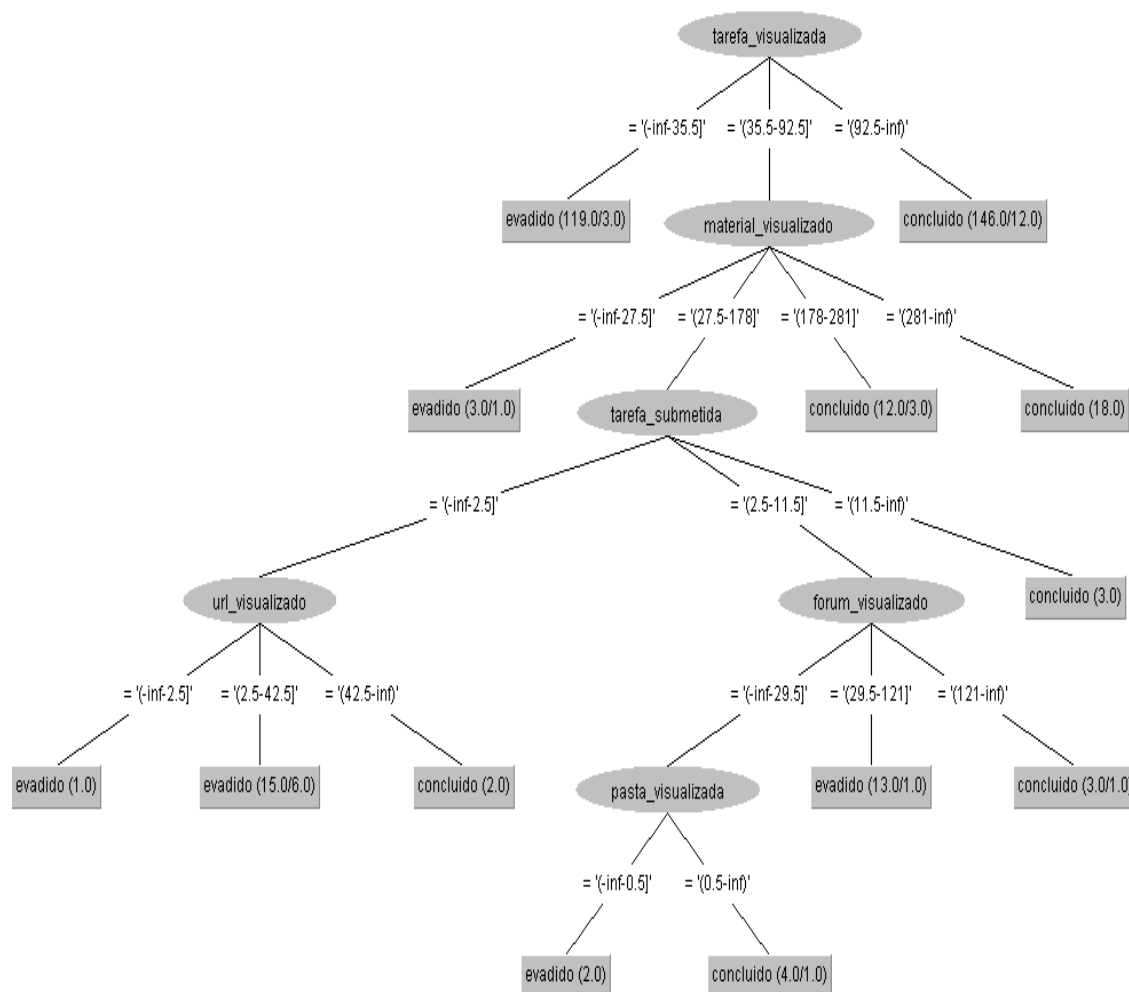
Fonte: Autor (2020).

Analisando a figura 10, percebe-se que os recursos que classificam melhor o aluno são a tarefa visualizada e o material visualizado. Os alunos que visualizaram 6% ou menos do total de interações no atributo tarefa visualizada, evadiram. Para quem interagiu 7%, ou mais, o recurso a ser analisado pelo algoritmo foi o material\_visualizado. Alunos que interagiram mais que 15% do número máximo de materiais visualizados, concluíram o curso, e, para os outros que interagiram 15% ou menos, o recurso a ser analisado é a visualização de material. Após se analisar em dois nós da árvore o recurso visualização de material, novamente é analisado o recurso tarefa visualizada, indicando que alunos que interagiram 19% nas tarefas, ou mais, concluíram o curso. Para os outros que visualizaram menos, o recurso tarefa submetida passa a ser analisado.

No terceiro experimento, por possuir um conjunto maior de dados, foi possível abordar uma outra forma de apresentar os dados na árvore de decisão, ou

seja, com intervalos de valores. Esses intervalos são definidos automaticamente pela ferramenta WEKA ao selecionar o filtro *discretize*, abordado na seção 3.1.2 deste trabalho, também chamado de discretização. O resultado pode ser visualizado na figura 11.

Figura 11 – Árvore de decisão com os dados discretizados



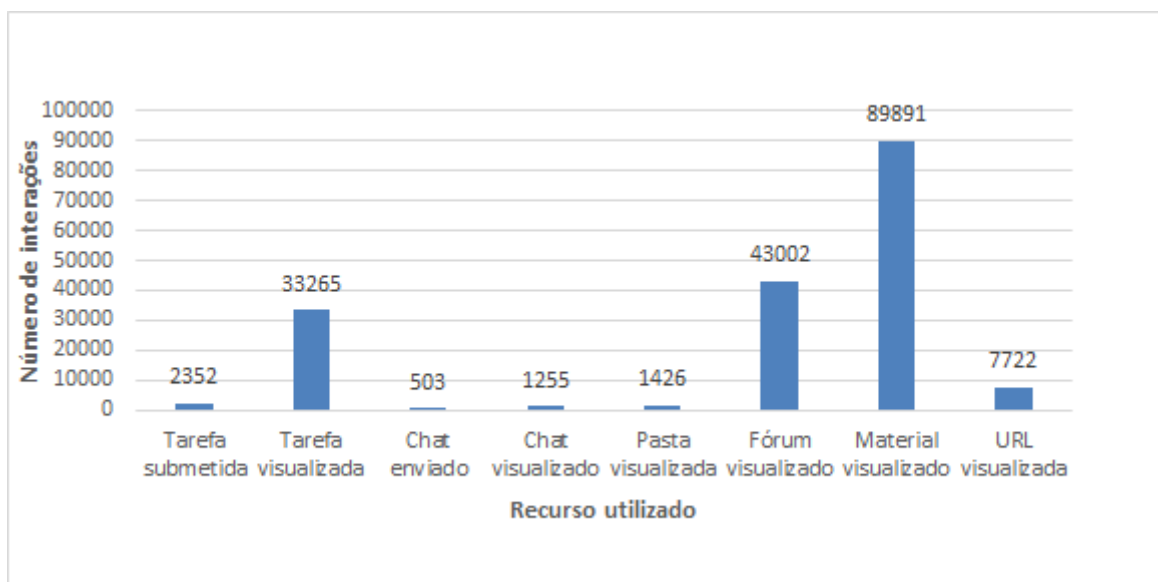
Fonte: Autor (2020)

Com os dados discretizados, a árvore de decisão mostra o recurso tarefa visualizada em 3 intervalos. De 0 a 35,5 interações, 119 alunos evadiram. No intervalo de 35,5 até 92,5 interações, o recurso visualização de material é analisado. Acima de 92,5 interações, 146 alunos foram classificados como concluído. Ao ser analisado o recurso material, no intervalo de 0 a 27,5 interações, 3 alunos foram classificados como evadido. No intervalo de 27,5 a 178, o recurso tarefa submetida é analisado. Quem visualizou mais de 178 vezes os materiais concluiu o curso,

sendo contabilizados 30 alunos. Quando analisado o recurso tarefa submetida pelo algoritmo, este verificou que, quando alguém interagiu até 2,5 vezes, houve a necessidade de o algoritmo analisar outro recurso; nesse caso, o recurso a ser analisado foi a visualização de *links*, ou seja, *url\_viewed*. No intervalo de 2,5 a 11,5, é analisado o recurso fórum. Quem submeteu mais de 11,5 tarefas concluiu o curso.

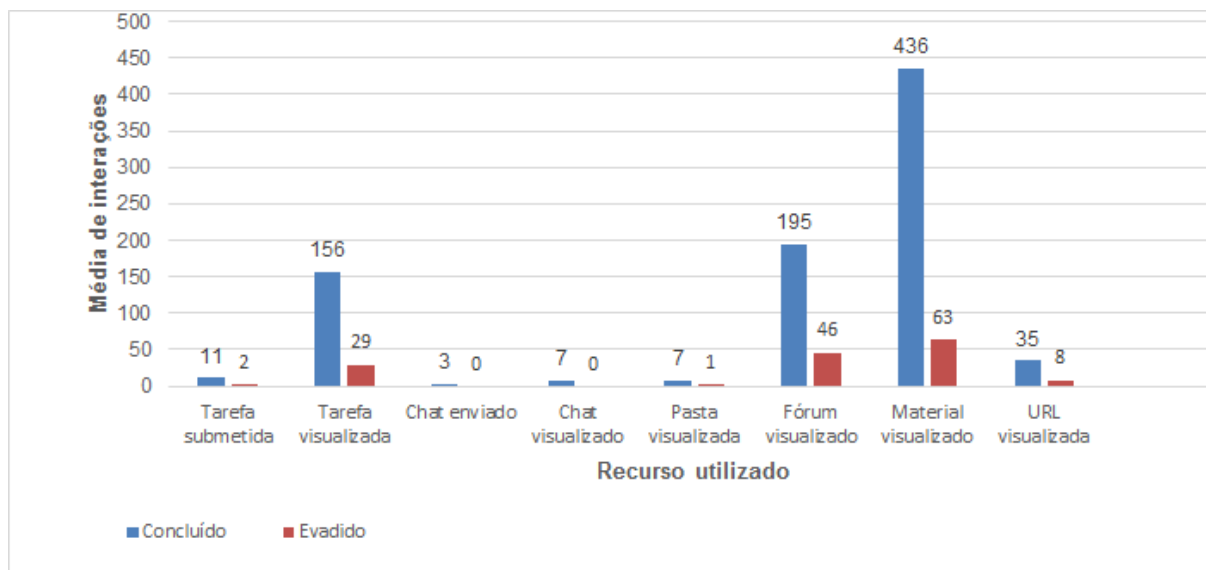
No gráfico 7, podemos visualizar o número total de interações em cada recurso utilizado nesse conjunto.

Gráfico 7 – Interações totais nos recursos utilizados do AVEA: terceiro experimento



Fonte: Autor (2020).

Gráfico 8 – Média de utilização dos recursos no AVEA: terceiro experimento



Fonte: Autor (2020).

O gráfico 7 mostra os três recursos mais utilizados, material, fórum e tarefa, com 89.891, 43.002 e 33.265 visualizações, respectivamente. O recurso material é o mais utilizado, tendo em vista a principal utilização nos experimentos 1 e 2. Os recursos chat, tarefa submetida, pasta e URL visualizada foram os menos utilizados.

No gráfico 8, como nos experimentos anteriores, podemos observar a média de utilização dos recursos nas duas classes, considerando 182 alunos concluintes e 159 evadidos.

Por refletir os mesmos dados em um único conjunto, a média de utilização segue a mesma proporção dos experimentos 1 e 2. O recurso material visualizado teve uma média de 436 interações pelos alunos concluintes e uma média de 63 interações pelos alunos evadidos. Média de 195 interações na classe de alunos concluintes e 46 para os evadidos no recurso fórum. Observa-se, também, a média de utilização no recurso chat e pasta visualizada para os alunos evadidos, sendo 0 interação no recurso chat, ou seja, não enviaram nem visualizaram o recurso, e uma interação apenas para o recurso pasta.

### 6.3.2 Análise dos dados: terceiro experimento

Ao analisar os dados, observam-se alguns fatores. Nem sempre a

quantidade de acessos ao atributo define a permanência do estudante no curso. Observa-se a árvore de decisão da figura 10, em que alguns alunos visualizaram o recurso mais vezes, porém evadiram. Pode-se deduzir que isso é em virtude de o aluno apenas visualizar o material e não atentar em realizar a tarefa, ou não se sentir apto para aprender a distância.

A interação com o aluno, o diálogo personalizado e a atenção ao sentimento de pertencimento do estudante não estão explícitos nas disciplinas, a partir dos dados que possuíamos. O estudante a distância deve ser estimulado constantemente para a realização das tarefas, esclarecer dúvidas nas disciplinas, sentir-se pertencente e identificado com a instituição, essa é uma interação que não visualizamos pelas árvores e que se faz indispensável para a permanência e o êxito dos estudantes.

Alguns autores propõem medidas para minimizar a evasão, como: a abordagem centrada no aluno, que “exige metodologias ativas de sala de aula que envolvam o aluno no processo de aprendizagem e que dependam da entrada do estudante para dar sentido aos objetivos institucionais” (DIAZ; BONTENBAL, 2001); e a construção de comunidade de aprendizagem, para que os alunos trabalhem juntos e ampliem sua base de conhecimentos de forma colaborativa (ANDERSON, 2004).

Outro fator importante é a diferença na utilização dos recursos entre os dois cursos. No experimento 1, a turma de Administração utilizou o recurso chat; já no experimento 2, o curso de Agroindústria não utilizou. Também, nota-se outra diferença no recurso fórum. No experimento 1, houve 462 interações, já no experimento 2, 1.639. Essa diferença pode ser explicada justamente pelo fato de o experimento 2 não utilizar o recurso chat, sendo, então, utilizado o fórum pela turma para discussões acerca do tema, ou para avisos gerais das disciplinas. Nesse cenário, observamos também que existe uma cultura do *campus* em relação aos usos do AVEA Moodle, pois o curso e suas diversas disciplinas possuem um padrão de organização das salas virtuais que difere entre os dois *campi*.

Ainda, sobre a frequência nos fóruns e chats, destaca-se que a presença virtual do estudante é definidora no processo de ensino-aprendizagem, mas ela está relacionada à presença virtual e interação do docente e do tutor no sentido de acolher e provocar o estudante. Além disso, em determinados grupos de estudantes, as barreiras tecnológicas e de tempo para a dedicação à rotina de

estudos faz com que esse estudante não participe de atividades síncronas, ou não busque esclarecer suas dúvidas, o que justifica a pouca utilização do chat e, até mesmo, do fórum.

Sendo assim, a qualificação e capacitação dos docentes e tutores é fundamental para a transposição didática dos conteúdos e construção de identidade no trabalho docente virtual. Somado a isso, conhecer a realidade do estudante, seu contexto social, sua fluência tecnológica e suas rotinas de trabalho e estudo auxilia o polo de apoio presencial e os tutores e professores a buscarem subsídios de inserção desses alunos no curso e na instituição. Essas questões podem ser inicialmente averiguadas a partir de extrações e análises de dados, como as deste estudo.

#### 6.4 ANÁLISE DA MINERAÇÃO ENTRE OS EXPERIMENTOS

Ao analisar os três experimentos, podemos perceber a diferença entre as métricas de avaliação. Para o experimento 1, o algoritmo com maior taxa de acerto foi o *Randon Forest*, com 93%.

No experimento 2, o algoritmo com maior acurácia foi o *J48*, com 85,50%. E, ao analisar o terceiro experimento que uniu os dois conjuntos, há em destaque o algoritmo *Randon Forest*, com taxa superior a 88%.

De forma geral, os três experimentos possuem boa taxa de acerto e a diferença deve-se à forma de utilização e a características dos alunos nos dois cursos. Alguns alunos acabam interagindo pouco com os recursos e, por fim, acabam concluindo o curso. O contrário também ocorre, alunos interagem mais, porém acabam evadindo. O algoritmo classifica o aluno buscando um padrão de utilização, sendo assim, a classificação atribuída é baseada em seu comportamento dentro do AVEA, ou seja, em suas interações com os recursos.

Esse comportamento fica mais evidente no experimento 2, pois a taxa de acerto foi um pouco menor, e o número de alunos matriculados foi menor, se comparado com o experimento 1. O algoritmo errou mais com um menor número de alunos.

Com isso, o experimento 3 mostra a sua importância, unindo os conjuntos de dados de forma a obter um maior número de alunos com características diferentes em um único *dataset*. Isso nos permite tentar buscar um padrão de evasão mais

apropriado e genérico.

Nesse mesmo viés, temos uma árvore com os dados discretizados no experimento 3, sendo que consideramos essa árvore a que melhor caracteriza e classifica os alunos pelo número de interações em cada recurso. Isso porque ela indica intervalos predefinidos pela ferramenta WEKA, o que nos permite visualizar os intervalos de interações em que o aluno foi classificado.

Por fim, devemos atentar para a diferença entre as médias de utilização dos recursos utilizados pelos alunos concluintes e evadidos. Ao analisar os atributos com maior êxito na classificação pelos algoritmos, identificamos que a média de utilização pelos alunos concluintes do atributo material visualizado foi de 436 interações durante todo o semestre, e, para os alunos evadidos, a média de utilização foi de 63 interações. A mesma situação ocorre ao analisar o atributo tarefa visualizada, para o qual a média de utilização pelos alunos concluintes foi de 156 interações, e a média de utilização pelos alunos evadidos, de 29 interações. Isso mostra a disparidade que há entre a utilização desses recursos pelas duas classes de alunos e deixa em evidência a característica do aluno evadido.



## 7 CONCLUSÃO

Neste trabalho, foi possível compreender a relevância do tema evasão, bem como as discussões sobre suas principais causas e abordagens, por diferentes autores na literatura. A permanência e o êxito dos estudantes são um tema importante para qualquer instituição de ensino e são do interesse dos gestores educacionais nos diferentes níveis que atuam; por isso, observar fatores relacionados à evasão faz-se tão importante.

Na educação a distância, por meio do AVEA Moodle, temos uma ferramenta que precisa ser explorada em seus mais diversos recursos, de forma que possamos utilizar melhor sua capacidade e, também, potencializar a mediação didática, qualificando a interação entre aluno, tutor, conteúdo e professor.

Deste modo, este trabalho buscou, por meio da mineração de dados educacionais e da interação do aluno no ambiente Moodle, dados e indicadores que pudessem ajudar os gestores a conhecer melhor a usabilidade da ferramenta, as características dos alunos com tendência a evadir, auxiliando na tomada de decisão da instituição. Essas decisões envolvem qualificação da equipe multidisciplinar e acompanhamento da frequência virtual dos estudantes para além do acesso e observação dos fatores socioculturais e experienciais dos ingressantes.

Foram realizados três experimentos, utilizando cursos distintos, ofertados por *campi* diferentes. O terceiro experimento consistiu na unificação dos dados para aumentar as características dos dois cursos em uma única base de dados, buscando o maior número de interações possíveis para a mineração. Em cada experimento, foi gerada a árvore de decisão em porcentagem. Para o terceiro experimento, foi possível uma segunda árvore, com os dados discretizados, ou seja, com os intervalos de interações em determinadas faixas, classificando se o aluno evadiu ou concluiu o curso considerando determinado recurso. Essas árvores, em estilos distintos, possibilitam uma leitura ampla sobre as interações em diferentes aspectos, conforme pode ser observado nas análises qualitativas.

Ainda, sobre o experimento 3, este foi considerado o que melhor apresenta os dados de evasão e mineração, sendo o experimento que consistiu na união do experimento 1 e 2, trazendo mais características dos alunos e um maior conjunto de dados para a realização da mineração.

No entanto, foi verificada pouca utilização dos recursos do Moodle pelos tutores e professores, limitando a utilização de mais recursos na mineração. Outro fator limitante foi a impossibilidade de utilizar a mineração dos dados dos cursos ofertados no ano de 2017-2018, devido ao fato de os dados de evasão não serem claros nos dados disponibilizados pelo IFFar.

Sendo assim, os dados da mineração que uniu as interações das duas turmas, realizada no experimento 3, mostram uma taxa de acerto superior a 88% com o algoritmo *Randon Forest*. Os melhores atributos que realizaram a predição foram a tarefa\_visualizada e resource\_viewed, ou seja, os indicadores de evasão para os cursos minerados foram a visualização de tarefas e de materiais disponibilizados. Isso leva à dedução do perfil do aluno concluinte nesses cursos minerados, ficando este submetido a apenas materiais e tarefas disponibilizadas, buscando responder às tarefas propostas pelo tutor e visualizando os materiais disponibilizados. Já o perfil do aluno evadido pode ser entendido como o aluno que pouco interage com os recursos do Moodle e que não visualiza os materiais e as tarefas disponibilizadas.

Por fim, com a análise qualitativa, foi possível ter uma melhor interpretação dos dados; também, foi de fundamental importância para entender as características de cada experimento analisado, já que nenhum outro estudo dessa natureza foi realizado, até então, no IFFar.

Ainda, em relação aos trabalhos correlatos relacionados nesta pesquisa, esta traz a vantagem de realizar uma análise qualitativa junto aos dados da mineração, abordando os resultados de forma clara, objetiva, e buscando entender os dados quantitativos.

Com esses dados encontrados, o IFFar poderá criar ações para melhorar a usabilidade do AVEA pelos tutores e alunos, criando métodos pedagógicos mais eficientes, capazes de incentivar o aluno de modo que não abandone o curso no qual está matriculado.

Como trabalhos futuros, sugere-se a utilização da mineração de dados em outros cursos EaD da instituição e, da mesma forma, em outros níveis de ensino. Também se sugere o emprego de outras informações existentes no AVEA, tais como a utilização do quiz, que não foi possível neste trabalho, interações pelos tutores, dados relacionados à localização dos polos de apoio e, se possível, as notas dos alunos.

Ainda, sugere-se a automatização dos resultados em uma ferramenta ou *plugin*, de forma que os gestores e tutores possam acompanhar frequentemente as interações do aluno e potencializar a permanência e o êxito do estudante a distância.

## REFERÊNCIAS

- ABBAD, G.; CARVALHO, R. S.; ZERBINI, T. Evasão em curso via internet: explorando variáveis explicativas. **RAE-eletrônica**, v. 5, n. 2, jul./dez. 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_abstract&pid=S1676-56482006000200008&lng=en&nrm=iso&tlng=pt](http://www.scielo.br/scielo.php?script=sci_abstract&pid=S1676-56482006000200008&lng=en&nrm=iso&tlng=pt)>. Acesso em: 10 dez. 2009.
- AGRAWAL, R.; IMIELIŃSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. **Acm sigmod record**. ACM, v. 22, n. 2, p. 207-216, 1993.
- ALMEIDA, M. E. B. de. Educação a distância na internet: abordagens e contribuições dos ambientes digitais de aprendizagem. **Educ. Pesqui.**, v. 29, n. 2, p. 327-340, 2003. Disponível em: <[http://www.scielo.br/scielo.php?pid=S1517-97022003000200010&script=sci\\_abstract&tlng=pt](http://www.scielo.br/scielo.php?pid=S1517-97022003000200010&script=sci_abstract&tlng=pt)>. Acesso em: 4 jul. 2019.
- ALMEIDA, O. C de S. de et al. Evasão em cursos a distância: fatores influenciadores. **Revista Brasileira de Orientação Profissional**, v. 14, n. 1, jan./jun., p. 19-33, 2013.
- ANDERSON, T. Toward a theory of online learning. In: ANDERSON, T.; ELLOUMI, F. (Org.). **Theory and practice of online learning**. Athabasca: Athabasca University, 2004. p. 33-60.
- ANDRADE, F. Educação a distância x Educação Presencial: algumas diferenças encontradas. **Blog Artigonal**, 2010. Disponível em: <<https://administradores.com.br/artigos/educacao-a-distancia-x-educacao-presencial-algumas-diferencas-encontradas>>. Acesso em: 4 jul. 2019.
- ASSOCIAÇÃO BRASILEIRA DE EDUCAÇÃO A DISTÂNCIA. **Relatório Analítico da Aprendizagem a Distância no Brasil 2016**. Curitiba: InterSaberes, 2017.
- BAKER, R. S. J.; ISOTANI, S.; CARVALHO, A. M. J. B. Mineração de Dados Educacionais: Oportunidades para o Brasil. **Revista Brasileira de Informática na Educação**, v. 19, n. 2., 2011. Disponível em: <<http://www.brie.org/pub/index.php/rbie/article/view/1301/1172>>. Acesso em: 4 jul. 2019.
- BARROSO M. F.; FALCÃO, E. B. M. Evasão universitária: o caso do Instituto de Física da UFRJ. In: ENCONTRO NACIONAL DE PESQUISA EM ENSINO DE FÍSICA, 9., 2004, Jaboticatubas. **Anais...** Jaboticatubas: Sociedade Brasileira de Física, 2004. p. 1-14.
- BELLONI, M. L. **Educação a distância**. 2. ed. Campinas, SP: Autores Associados, 2001.
- BITTENCOURT, I. M.; MERCADO, L. P. L. Evasão nos cursos na modalidade de educação a distância: estudo de caso do Curso Piloto de Administração da

UFAL/UAB. **Revista Ensaio: Avaliação de Políticas Públicas em Educação**. Rio de Janeiro, v. 22, n. 83, p. 465-504, abr./jun. 2014.

BIZARRIA, F. P. A.; SILVA, M. A.; CARNEIRO, T. C. J. Evasão discente na EAD: percepções do papel do tutor em uma instituição de ensino superior. In: CONGRESSO BRASILEIRO DE ENSINO SUPERIOR A DISTÂNCIA – ESUD. 11., 2014, Florianópolis. **Anais...** Florianópolis: UFSC, 2014. Disponível em: <<http://esud2014.nute.ufsc.br/anais-esud2014/>>. Acesso em: 20 set. 2016.

BRACHMAN, R. et al. Mining business databases. **Communications of the ACM**, v. 39, n. 11, p. 42-48, 1996.

BRASIL. MEC, SETEC, IFFAR, **PDI 2019-2026**. Disponível em: <https://www.iffarroupilha.edu.br/documentos-do-pdi/item/13876-pdi-2019-2026>

BURGOS, C. et al. Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. **Computers & Electrical Engineering**. p. 1-16, mar. 2017.

CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados**: conceitos básicos, algoritmos e aplicações. 1. ed. São Paulo: Saraiva, 2016.

COELHO, M. L. A. **Evasão nos Cursos de Formação Continuada de Professores Universitários na Modalidade de Educação a Distância Via Internet** - Universidade Federal de Minas Gerais, 2002.

COSTA, E. et al. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. **Jornada de Atualização em Informática na Educação**. 2012.

COSTA, E. et al. Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. **Jornada de Atualização em Informática na Educação**, v. 1, n. 1, p. 1-29, 2013.

DA COSTA, S. S.; CAZELLA, S.; RIGO, S. J. Minerando Dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EaD: Um estudo de caso sobre evasão escolar na UNA-SUS. **RENOTE**, v. 12, n. 2, 2014.

DETONI, D.; ARAÚJO, R.; CECHINEL, C. Modelling and Prediction of Distance Learning Students Failure by using the Count of Interactions. **Revista Brasileira de Informática na Educação**, v. 23, p. 1-11, 2015.

DIAZ, D. P.; BONTENBAL, K. F. Learner preferences: Developing a learner-centered environment in the online or mediated classroom. **Education at a Distance**, v. 15, n. 8, 2001.

DORE, R.; LÜSCHER, A. Z. Permanência e evasão na educação técnica de nível médio em Minas Gerais. **Cadernos de Pesquisa**, v. 41, n. 144, p. 772-789, 2011.

DOUGHERTY, J. et al. **Supervised and unsupervised discretization of continuous features**, *Machine learning: proceedings of the twelfth international conference*, v. 12, p. 194–202, 1995.

EYNG, A. M.; GISI, M.; ENS, R.; PACIEVITCH, T. Diversidade e padronização nas políticas educacionais: configurações da convivência escolar. **Ensaio: Avaliação e Políticas Públicas em Educação**, [S.l.], v. 21, n. 81, p. 773–800, 2013.

FAVERO, R. V. **Dialogar ou evadir: eis a questão**: um estudo sobre a permanência e a evasão na educação a distância no estado do Rio Grande do Sul. 2006. Dissertação (Mestrado) – Programa de Educação a Distância, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**. Providence, v. 17, n. 3, p. 37-54, jul. 1996.

FIUZA, P. J. **Adesão e permanência discente na Educação à distância: investigação de motivos e análise de preditores sociodemográficos, motivacionais e de personalidade para o desempenho na modalidade**. Porto Alegre, RS. 2012. 145p. Tese (Doutorado) – Universidade Federal do Rio Grande do Sul, 2012.

GALVÃO, N. D. **Técnica de mineração de dados: uma revisão da literatura**. Cuiabá: Editora, 2009.

GARCIA, A. C. **Mineração de dados aplicada a sistemas de recomendação**. Trabalho de Conclusão de Curso (Graduação) – Universidade de Santa Cruz do Sul, Santa Cruz do Sul, 2012.

GARCIA, S. et al. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning, *Knowledge and Data Engineering*. **IEEE Transactions**, v. 25, n. 4, p. 734–750, 2013.

GOLDSCHMIDT, R.; BEZERRA, E.; PASSOS, E. **Data mining**: conceitos, técnicas, algoritmos orientações e aplicações. 2. ed. Rio de Janeiro: Elsevier, 2015.

HAN, J.; KAMBER, M. **Data Mining**: Concepts and Techniques. Morgan Kaufmann, 2001.

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

KENSKI, V. M. **Educação e tecnologias**: o novo ritmo da informação. Campinas, SP: Papyrus, 2007.

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

MAIA, C.; J. MATTAR. **ABC da EaD: a Educação a Distância hoje**. 1. ed. São Paulo: Pearson. 2007.

MANHÃES, L. M. B. **Predição do desempenho acadêmico de graduandos utilizando mineração de dados educacionais**. 2015. 157 p. Tese (Doutorado em Engenharia de Sistemas e Computação) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Rio de Janeiro, 2015.

MARTÍNEZ, M.; GARCÍA M. C.; MONTORO, J. M. **Dificuldades de aprendizagem**. 1. ed. Porto: Porto editora, 2003.

MEC. SETEC. IFFARROUPILHA. **Programa Permanência e Êxito**. Instituto Federal Farroupilha:2014. Disponível em: <<http://w2.iffarroupilha.edu.br/site/conteudo.php?cat=168&sub=6013>>. Acesso em: 4 nov. 2019.

MILL, D. et al. Gestão da Educação a Distância (EaD): noções sobre planejamento, organização, direção e controle da EaD. **Vertentes (UFSJ)**, v. 35, p. 9-23, 2010.

MILL, D. Mudanças de mentalidade sobre educação e tecnologia: inovações possibilidades tecnopedagógicas. In: MILL, D. (Org.) **Escritos sobre educação: desafios e possibilidades para ensinar e aprender com as tecnologias emergentes**. São Paulo: Paulus, 2013.

MILL, D. Sobre o conceito de polidocência ou sobre a natureza do processo de trabalho pedagógico na educação a distância. In: MILL, D; RIBEIRO, L. R. de C.; OLIVEIRA, M. R. G. de (Org.). **Polidocência na educação a distância: múltiplos enfoques**. São Carlos: EdUFSCar, 2014.

MITCHELL, T. M. Machine learning and data mining. **Communications of the ACM**, v. 42, n. 11, p. 30-36, 1999.

MORAN, J. M. **A educação que desejamos: novos desafios e como chegar lá**. Campinas, SP: Papirus, 2007.

MORAN, J. M. O que é Educação a Distância. Universidade de São Paulo. 2002. Disponível em: <<http://www2.eca.usp.br/moran/wp-content/uploads/2013/12/dist.pdf>>. Acesso em: 18 mar. 2019.

OLIVEIRA, E. H. T. et al. Distance Education with remote poles: an example from the Amazon region. In: **Frontiers in Education (FIE)**. Seattle, WA: Editora, 2012.

OLIVEIRA, F. B. **considerações sobre educação a distância no ensino superior: a experiência da Fundação Getulio Vargas**. Rio de Janeiro: Editora, 2009.

OLIVEIRA, J. J. G.; NORONHA, R. V.; KAESTNER, C. A. A. Método de seleção de atributos aplicados na previsão da evasão de cursos de graduação. **Revista de Informática Aplicada**. 2017.

PEREIRA, F. C. B. **Determinantes da evasão de alunos e os custos ocultos para as instituições de ensino superior**: uma aplicação na universidade do extremo sul catarinense. Tese (Doutorado) – Universidade Federal de Santa Catarina, Florianópolis, 2003.

QUEIROGA, E.; CECHINEL, C.; ARAÚJO, R. Predição de estudantes com risco de evasão em cursos técnicos a distância. In: **Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)**. Recife: Sociedade Brasileira de Computação, 2017. v. 1, p. 1547-1556.

RABELO, H. et al. Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EaD em ambientes virtuais de aprendizagem. **Anais do SBIE 2017**, p. 1527-1536, 2017.

RAMOS, W. M. Evasão em cursos a distância: fatores intervenientes. In: **III Jornada em Educação a Distância em Letras – Português UFSC**. Coordenação EAD realizada na Universidade Federal de Santa Catarina. 2014. Disponível em: <<https://uab.ufsc.br/portugues/files/2014/03/Fen%c3%b4meno-da-Evas%c3%a3o-e-da-Persist%c3%aancia-conceitual-vfinal-ufsc.pdf>>. Acesso em: 3 jan. 2015.

RAMOS, W. M.; BICALHO, R. N. M; SOUSA, J. V. de S. **Evasão e persistência em cursos superiores a distância**: o estado da arte da literatura internacional. Portugal: Universidade de Coimbra, 2015.

RIGO, S. J. et al. Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação**. v. 22, n. 1, 2014.

RIGO, S. J.; CAZELLA, S. C.; CAMBRUZZI, W. Minerando Dados Educacionais com foco na evasão escolar: oportunidades, desafios e necessidades. **Anais do Workshop de Desafios da Computação Aplicada à Educação**, p. 168-177, 2012.

RODRIGUES, R. L.; DE MEDEIROS, F. P.; GOMES, A. S. Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem. In: **Anais do XXIV Simpósio Brasileiro de Informática na Educação**. 2013.

ROMERO, C.; VENTURA, S. Data mining in education. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 3, n. 1, p. 12-27, 2013.

SANCHEZ, F. (Coord.). **Anuário Brasileiro Estatístico de Educação Aberta e a Distância - ABRAEAD**. 4. ed. São Paulo: Instituto Monitor, 2008.



SANTANA, L. C.; MACIEL, A. M.; RODRIGUES, R. L. Avaliação do perfil de uso no ambiente Moodle utilizando técnicas de mineração de dados. In: **Anais do Simpósio Brasileiro de Informática na Educação**. [S.l.: s.n.], v. 25, n. 1, p. 269, 2014.

SANTOS, E. O. dos. Articulação de saberes na EAD online: por uma rede interdisciplinar e interativa de conhecimentos em ambientes virtuais de aprendizagem. In: SILVA, M. (Org.). **Educação online: teorias, práticas, legislação, formação corporativa**. São Paulo: Edições Loyola, 2003. p. 217-230.

SCHMITT, J. A. **Identificação de alunos com tendência a evasão nos cursos de graduação a distância por meio de mineração de dados educacionais**. 2018. 175 p. Dissertação (Mestrado Profissional em Tecnologias Educacionais em Rede) – Universidade Federal de Santa Maria, Santa Maria, 2018.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: From theory to algorithms**. [S.l.]: Cambridge university press, 2014.

SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistema de Banco de Dados**. Tradução de Daniel Vieira. Rio de Janeiro: Elsevier, 2006.

SILVA FILHO, R. L. L. et al. A evasão no Ensino Superior brasileiro. **Cadernos de Pesquisa**, São Paulo, v. 37, n. 132, p. 641-659, set./dez. 2007.

SILVA, D. G. **Análise sobre o uso dos relatórios de atividades do Moodle no acompanhamento do processo de aprendizagem de alunos em cursos de graduação**. Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Mato Grosso, 2011.

SILVA, R. S. **Moodle para autores e tutores: educação a distância na web 2.0**. 1. ed. São Paulo: Novatec, 2010.

SIMPSON, O.; WOODLEY, A. Evasão: o elefante na sala. In: ZAWACKI-RICHTER, O.; ANDERSON, T. **Educação a distância online: construindo uma agenda de pesquisa**. Tradução de Isabela de Martini Rivera Ferreira. São Paulo: Artesanato Educacional, 2015.

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

TOCZEK, J. et al. **Uma visão macroscópica da evasão no ensino superior a distância do Brasil**. Disponível em: . Acesso em: 17 nov. 2019.

WEBBER, C. G.; ZAT, D.; LIMA, M. F. W. P. Utilização de algoritmos de agrupamento na mineração de dados educacionais. **Revista Renote: Novas tecnologias na educação**, v. 11, n. 1, 2013.

WEKA, 2019 UNIVERSITY OF WAIKATO. Weka 3.8 – Machine Learning Software in Java. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/downloading.html>>. Acesso em: 18 nov. 2019

WITTEN, I. H.; FRANK, E. **Data Mining**: Practical Machine Learning Tools and Techniques. São Francisco: Morgan Kaufmann Publishers, 2005.

ZAKI, M. J. **Parallel and Distributed Data Mining**: An Introduction. Large-Scale Parallel Data Mining. Berlin: Springer-Verlag, 2000.

## ANEXO A – SOLICITAÇÃO



Ministério da Educação  
Universidade Federal de Santa Maria  
Centro de Educação  
PPGTER - Programa de Pós-Graduação em  
Tecnologias Educacionais em Rede  
**MPTER – Mestrado Profissional em  
Tecnologias Educacionais em Rede**



CE  
CENTRO DE EDUCAÇÃO  
UFSM

### CARTA DE APRESENTAÇÃO

Sra. Carla Comerlato Jardim  
Magnífica Reitora  
Instituto Federal Farroupilha

De acordo  
Angela M. A. Marinho  
Chefe de Gabinete  
Instituto Federal Farroupilha  
Portaria 927/2012

Prezado Reitora,

Ao cumprimentá-la cordialmente, vimos pelo presente apresentar e solicitar autorização para realização do projeto de pesquisa de mestrado intitulado “UTILIZAÇÃO DE MINERAÇÃO DE DADOS EDUCACIONAIS PARA COMPOR INDICADORES DA TENDÊNCIA DE EVASÃO ESCOLAR EM CURSOS DE EAD NO INSTITUTO FEDERAL FARROUPILHA”, vinculado ao Programa de Mestrado em Tecnologias Educacionais em Rede, da UFSM e tendo como pesquisadores Solange Pertile (orientadora) e Thiago Siqueira Sonnenstrahl (estudante e servidor do IFFAR).

A pesquisa tem como objetivo geral, por meio da Mineração de Dados Educacionais (MDE), propor e avaliar indicadores da tendência de evasão em cursos do Instituto Federal Farroupilha na modalidade à distância, disponibilizando dados estratégicos para os gestores educacionais da instituição. e, como objetivos específicos Estruturar indicadores para identificação do perfil de alunos com tendência a evasão em cursos EaD do IFFar; Avaliar técnicas de MDE que auxiliem na análise preditiva de evasão, a fim de potencializar a permanência e êxito dos estudantes de cursos EaD e disponibilizar em uma ferramenta baseada na MDE com indicadores para identificação do perfil de alunos com tendência a evasão para gestores educacionais do IFFar. A pesquisa adotará a metodologia de pesquisa exploratória, baseada na mineração de dados.

Para que a pesquisa seja realizada, será NECESSÁRIO ACESSO (CÓPIA DA BASE) PARA VISUALIZAÇÃO DA BASE DE DADOS DO AMBIENTE VIRTUAL DE APRENDIZAGEM (Moodle) do IFFAR, uma vez que serão necessárias consultas aos cursos ofertados pela instituição na modalidade EAD bem como aos dados de acesso e uso, e de desempenho acadêmico dos estudantes na plataforma (os nomes dos alunos poderão ser ocultos, pois, não há necessidade de identificação destes). Ressalte-se que não será necessário o acesso aos dados pessoais dos estudantes, uma vez que o estudo busca apenas identificar e fornecer indicadores que possam contribuir com a diminuição da evasão escolar em cursos na modalidade EAD, e que todos os cuidados serão tomados para garantir o sigilo e a confidencialidade das informações, preservando a identidade dos estudantes/utilizadores da plataforma online.

Ao final da pesquisa, os resultados serão disponibilizados ao IFFAR e apresentadas/publicizadas com a comunidade acadêmica visando contribuir com o desenvolvimento científico/tecnológico/educacional.

Agradeço a disponibilidade e colaboração.

Atenciosamente,

Solange Pertile  
Docente PPGTER – UFSM

Thiago Siqueira Sonnenstrahl  
Aluno do PPGTER

## ANEXO B – AUTORIZAÇÃO

---



MINISTÉRIO DA EDUCAÇÃO  
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA  
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA FARROUPILHA  
REITORIA

### AUTORIZAÇÃO

Autorizamos o(a) servidor(a) THIAGO SIQUEIRA SONNENSTRAHL, Matrícula SIAPE N° 1800173, ocupante do cargo de Técnico-Administrativo em Educação – Técnico de Tecnologia da Informação, lotado(a) na Reitoria do Instituto Federal Farroupilha, a realizar o projeto de pesquisa de mestrado intitulado: “UTILIZAÇÃO DE MINERAÇÃO DE DADOS EDUCACIONAIS PARA COMPOR INDICADORES DA TENDÊNCIA DE EVASÃO ESCOLAR EM CURSOS DE EAD NO INSTITUTO FEDERAL FARROUPILHA”, vinculado ao Programa de Mestrado em Tecnologias Educacionais em Rede da Universidade Federal de Santa Maria (UFSM), tendo como orientadora Solange Pertile.

Santa Maria/RS, 12 de dezembro de 2018.

A handwritten signature in blue ink, appearing to read 'Carla Comerlato Jardim'.

CARLA COMERLATO JARDIM  
REITORA  
IF FARROUPILHA – RS