UNIVERSIDADE FEDERAL DO PAMPA

JULIANO MARCUZZO BORIN

DESENVOLVIMENTO DE UM SOTWARE PARA ANÁLISE DE EVASÃO NA UNIPAMPA CAMPUS BAGÉ UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Bagé
2014

JULIANO MARCUZZO BORIN

DESENVOLVIMENTO DE UM SOTWARE PARA ANÁLISE DE EVASÃO NA UNIPAMPA CAMPUS BAGÉ UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação da Universidade Federal do Pampa, como requisito para obtenção do Título de Bacharel em Engenharia da Computação.

Orientador: Sandro da Silva Camargo

Bagé
2014

JULIANO MARCUZZO BORIN

DESENVOLVIMENTO DE UM SOTWARE PARA ANÁLISE DE EVASÃO NA UNIPAMPA CAMPUS BAGÉ UTILIZANDO TÉCNICAS DE MINERAÇÃO DE DADOS

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação da Universidade Federal do Pampa, como requisito para obtenção do Título de Bacharel em Engenharia de Computação.

Trabalho de Conclusão de Curso defendido e aprovado em: 22 de março de 2014.

Banca examinadora:

_____

Prof. Dr. Sandro da Silva Camargo
Orientador
UNIPAMPA

_____

Profa. MSc. Sandra Dutra Piovesan
UNIPAMPA

_____

Prof. MSc. Gerson Alberto Leiria Nunes
UNIPAMPA

Dedico este trabalho à minha família, amigos, orientador e professores pelo apoio prestado durante o desenvolvimento do trabalho e minha graduação.

RESUMO

A evasão é um grande problema presente nas universidades, inclusive na Universidade Federal do Pampa, gerando prejuízos sociais, econômicos e acadêmicos. Estudar esse fenômeno e traçar metas eficazes para o combate à evasão é indispensável para uso adequado dos recursos das instituições de ensino. Porém, devido à elevada quantidade de dados disponíveis, o estudo de forma manual é proibitivo. O presente trabalho descreve o desenvolvimento de um *software* para processamento dos dados dos estudantes da UNIPAMPA Bagé e a aplicação de técnicas de Descoberta de Conhecimento em Bancos de Dados para a descoberta de padrões presentes entre os estudantes evadidos. São apresentados os estudos realizados com os dados dos alunos dos cursos de Engenharia de Computação, Engenharia de Alimentos, Engenharia de Produção, Engenharia Química, Engenharia de Energias Renováveis e de Ambiente e Licenciatura em Física, bem como os padrões obtidos para cada curso. Os padrões obtidos serão utilizados para a criação de mecanismos de combate à evasão na universidade, de forma que possa ocorrer um combate eficaz à evasão.

Palavras-Chave: evasão, desenvolvimento de *software,* Descoberta de Conhecimento em Bancos de Dados.

ABSTRACT

Dropping out is a big problem present in Universities, including in Universidade Federal do Pampa. This problem generates social, economical and academical losses. Studying this phenomena and tracing effective goals to avoid dropping out is indispensable to an appropriate use of resources of educational institutions. However, due to huge amount of available data, manual study is prohibitive. This work proposes the development of a software to process data of students from UNIPAMPA Bagé and the use of Knowledge Discovery in Databases techniques to discover patterns between data of students who dropped out. Studies realized with the data from the students of Computer Engineering, Food Engineering, Production Engineering, Chemical Engineering, Renewable Energy and Environment Engineering and Degree in Physics are listed. The patterns discovered will be used to create mechanisms to prevent dropping out in University, so that could occur an efficient prevention of dropping out.

Keywords: dropping out, software development, Knowledge Discovery in Databases.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

API – *Application Programming Interface* ou Interface de Programação de Aplicativos

CSV – *Comma Separated Values*

ENEM – Exame Nacional do Ensino Médio

IDE – *Integrated Development Enviroment* ou Ambiente Integrado de Desenvolvimento

KDD – *Knowledge Discovery in Databases* ou Descoberta de Conhecimento em Bancos de Dados

SIE – Sistema de Informações para Ensino

SiSU – Sistema de Seleção Unificado

UNIPAMPA – Universidade Federal do Pampa

WEKA – *Waikato Environment for Knowledge Analysis*

XLS – Formato de arquivos do Microsoft Excel[1]

[1] © Microsoft Corporation. Todos os direitos reservados.

SUMÁRIO

1 INTRODUÇÃO

A evasão no ensino superior brasileiro gera muitos prejuízos financeiros para o país, tendo valor estimado em torno de R$ 9 bilhões em 2009 [NOGUEIRA, 2011]. Cerca de 13,2% dos alunos de Universidades Federais evadiram em 2010 [BORGES, 2012]. Números como esses geram grande preocupação para todas as instituições de ensino superior, onde causam enormes perdas.

Tais perdas causam impacto negativo no desenvolvimento do país, pois as vagas deixadas por estudantes evadidos poderiam estar sendo utilizadas por outros que desejam estudar. Consequentemente, a evasão diminui o número de profissionais formados anualmente no Brasil [FILHO, 2007]. Tendo em vista a carência de profissionais qualificados no mercado, especialmente na área de engenharia, aumentam as consequências negativas geradas pela evasão [Falta ..., 2013].

O prejuízo também pode ser analisado pelo ponto de vista econômico para o país, onde existem receitas destinadas para manutenção de estruturas, professores e profissionais ligados às universidades, as quais não são aproveitadas em sua totalidade [FILHO, 2007]. Um melhor aproveitamento dos recursos dedicados ao ensino superior está ligado diretamente ao combate à evasão.

São muitos fatores que podem levar alunos a desistirem de seus estudos. Existem fatores desde dificuldades financeiras, de adaptação a uma nova cidade e a escolha de uma carreira a qual não atinge as expectativas criadas previamente [ALVES, 2001]. Essas situações não estão ligadas diretamente à vida acadêmica do aluno, sendo mais difícil determinar um padrão nos perfis dos evadidos.

Também existem fatores educacionais, como dificuldades de aprendizagem em determinados assuntos, professores pouco qualificados e reprovações [FILHO, 2007]. Essas informações podem fornecer possíveis perfis de alunos propensos à desistência. Entender e estudar quais fatores estão ligados à evasão é de vital importância para uma melhora da educação no Brasil.

Assim como ocorre em outras instituições, na Universidade Federal do Pampa a evasão também é um problema crítico. Sendo assim, é necessário fazer um levantamento dos perfis dos alunos desistentes e tentar buscar similaridades existentes entre eles. Isso possibilitará a adoção de medidas preventivas por partes

das coordenações e grupos de trabalho responsáveis pelo estudo e análise da evasão.

Porém, a utilização de um método manual para a analise dos perfis dos alunos evadidos é impraticável, tendo em vista o elevado número de estudantes inscritos na instituição. Dessa maneira, se torna imprescindível à utilização de um *software* que permita a automatização do processo, possibilitando rapidamente uma análise e tomadas de decisões necessárias.

1.1 Objetivos

Este trabalho descreve o desenvolvimento de um software que serve de auxílio ao estudo da evasão na UNIPAMPA, a fim de permitir a identificação de padrões nos dados dos alunos evadidos, permitindo a análise e desenvolvimento de medidas de combate. Os objetivos específicos deste trabalho foram:

- Obter os dados dos estudantes da UNIPAMPA Bagé;
- Desenvolver o *software* para automatização no pré-processamento de dados;
- Pré-processar os dados dos alunos;
- Minerar os dados;
- Pós-processar os dados;
- Analisar os resultados obtidos e gerar relatórios;
- Apresentar e discutir os resultados com os setores responsáveis pelo estudo da evasão.

1.2 Estrutura do Trabalho

Este trabalho está organizado da seguinte forma: o capitulo 2 descreve os conceitos necessários para embasamento teórico do trabalho; o capitulo 3 aborda os métodos adotados para o desenvolvimento do *software;* o capitulo 4 apresenta os resultados obtidos através da mineração dos dados gerados pela ferramenta; o capitulo 5 demonstra as considerações finais.

1.3 Trabalhos Anteriores

Este trabalho busca ampliar o estudo realizado por um trabalho anterior na própria universidade através de LANOT(2012). Nesse estudo as ferramentas

utilizadas para processamento dos dados exigiam que o utilizador possuísse um conhecimento das regras a serem utilizadas, pois a configuração dos parâmetros de processamento ocorria de forma manual, gerando uma limitação na quantidade de arquivos que foram manipulados durante o estudo. Com isso, a quantidade de cursos abrangidos foi pequena, pois o estudo foi dirigido apenas ao curso de Engenharia de Computação.

2 REVISÃO DE LITERATURA

Neste capítulo serão apresentadas as bases teóricas utilizadas para realização do trabalho. Serão descritos os conceitos de Evasão, Banco de Dados e Descoberta de Conhecimento em Banco de Dados. Após, serão descritos os principais passos desse processo, onde pode-se destacar a Mineração de Dados. Por último, será abordada a ferramenta WEKA.

2.1 Evasão

Evasão consiste no desligamento de um estudante de seu curso, sua instituição de ensino ou do Sistema Educacional. Caracteriza um dos maiores e mais preocupantes problemas do Sistema Educacional [MORAES, 2010]. Segundo FILHO(2007), as perdas causadas pelos evadidos geram desperdícios sociais, econômicos e acadêmicos. Essas perdas variam desde investimento de recursos sem o devido retorno até a ociosidade de professores, servidores, equipamentos e espaços físicos. Estratégias que buscam o combate à evasão requerem um elevado esforço por parte das instituições, pois é necessário conhecer os fatores que levam o estudante a abandonar seus estudos.

Fatores que ocasionam a evasão

A evasão não se caracteriza por um único fator apenas, podendo variar principalmente entre fatores econômicos e acadêmicos. As Instituições citam como principal fator a falta de recursos financeiros por parte do estudante para o prosseguimento dos estudos [FILHO, 2007]. Também ocorrem casos onde, para suprir a necessidade financeira, os estudantes enfrentam longas jornadas de trabalho, onde o cansaço acarreta na preferência pelos recursos financeiros aos estudos. Adicionalmente, merecem ser citados fatores como distância dos familiares e longas viagens para poder visitar sua cidade natal [MORAES, 2010].

Porém, existem fatores acadêmicos, como expectativas profissionais e dificuldades de aprendizagem. Os alunos estão acostumados com processos de ensino desde suas bases que não condizem com os adotados nas universidades. Muitos estão acostumados apenas com o processo de memorização, diferentemente

da necessidade de um espírito investigador presente no ensino superior. Também existem casos onde a escolha de profissão por parte do aluno foi equivocada, onde o erro somente é percebido durante o andamento da graduação [MORAES, 2010]. Outros consideram que o esforço necessário para obter o diploma não condiz com suas expectativas financeiras como profissional [FILHO, 2007].

Para descoberta desses fatores, uma alternativa é o estudo dos dados de cada estudante inscrito na universidade. Dados que são armazenados em estruturas destinadas para tal fim, os chamados Bancos de Dados.

2.2 Bancos de Dados

Bancos de Dados são basicamente sistemas computadorizados para armazenamento de registros, ou seja, seu propósito geral é armazenar dados e permitir aos usuários que acessem e atualizem os dados ali presentes quando necessário. As informações podem ser quaisquer que possuam importância para o individuo ou organização que utilize destas estruturas [DATE, 2003]. Como exemplos de bancos de dados, podem ser citados: estoque de produtos em um supermercado; controle do acervo de livros de uma biblioteca; ou relação de alunos matriculados em um curso [LANOT, 2012].

Dentre as principais funções de um banco de dados, devem ser permitidas a inclusão, armazenamento, manipulação e consulta de dados [DATE, 2003]. Em sua forma básica, os dados são organizados em Relações, Tuplas e Atributos:

- Relações são expressas na forma de tabelas, onde os dados ficam organizados em linhas e colunas. Também podem ser definidos como um conjunto de tuplas que possuem atributos em comum;
- Tuplas são as informações de determinado objeto presente na tabela. Como exemplo pode ser citado o número “20” como tupla para o atributo “Idade”;
- Atributos são as especificações de determinada coluna presente na tabela. Servem para expressar as características de determinado dado. Como exemplo, os campos “Cidade”, “Sexo” e “Nome” são atributos de uma tabela.

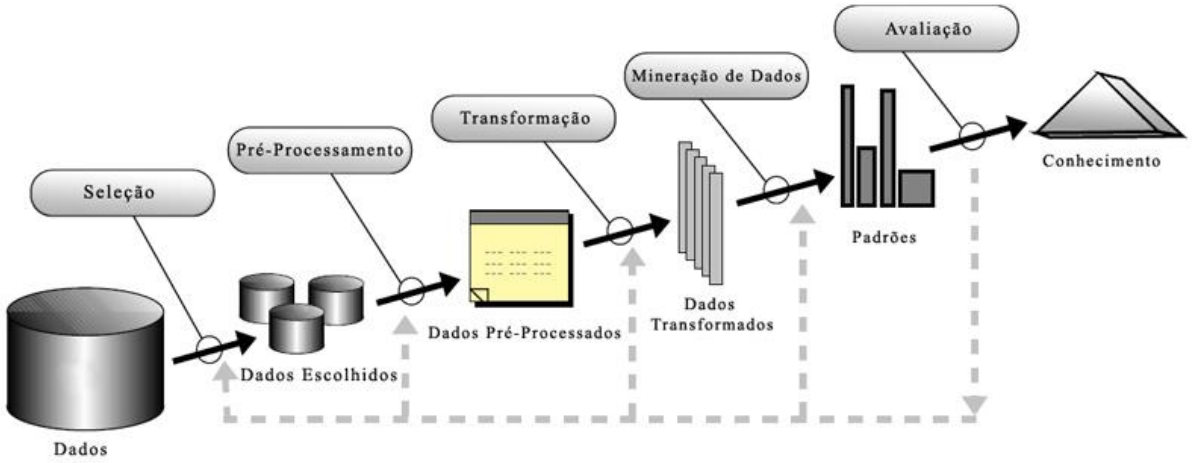
Considerando a grande quantidade de informação que pode ser armazenada em um banco de dados, a obtenção de padrões, comportamentos ou tendências expressos na forma de dados estatísticos torna-se possível através de um estudo

detalhado. Esse estudo é descrito através de técnicas chamadas de Descoberta de Conhecimento em Bancos de Dados [LANOT, 2012].

2.3 Descoberta de Conhecimento em Bancos de Dados

A maneira tradicional de obter conhecimento através de dados consiste na análise manual por especialistas das informações fornecidas pelos dados. Porém, em casos onde a quantidade de dados for elevada, o estudo manual se torna impraticável. Descoberta de Conhecimento em Bancos de Dados é uma tentativa de automatizar esse processo através de algoritmos que consigam extrair conhecimento com base em dados organizados previamente [CAMILO, 2009].

Figura 1 - Representação do processo de KDD (*Knowledge Discovery in Databases)*



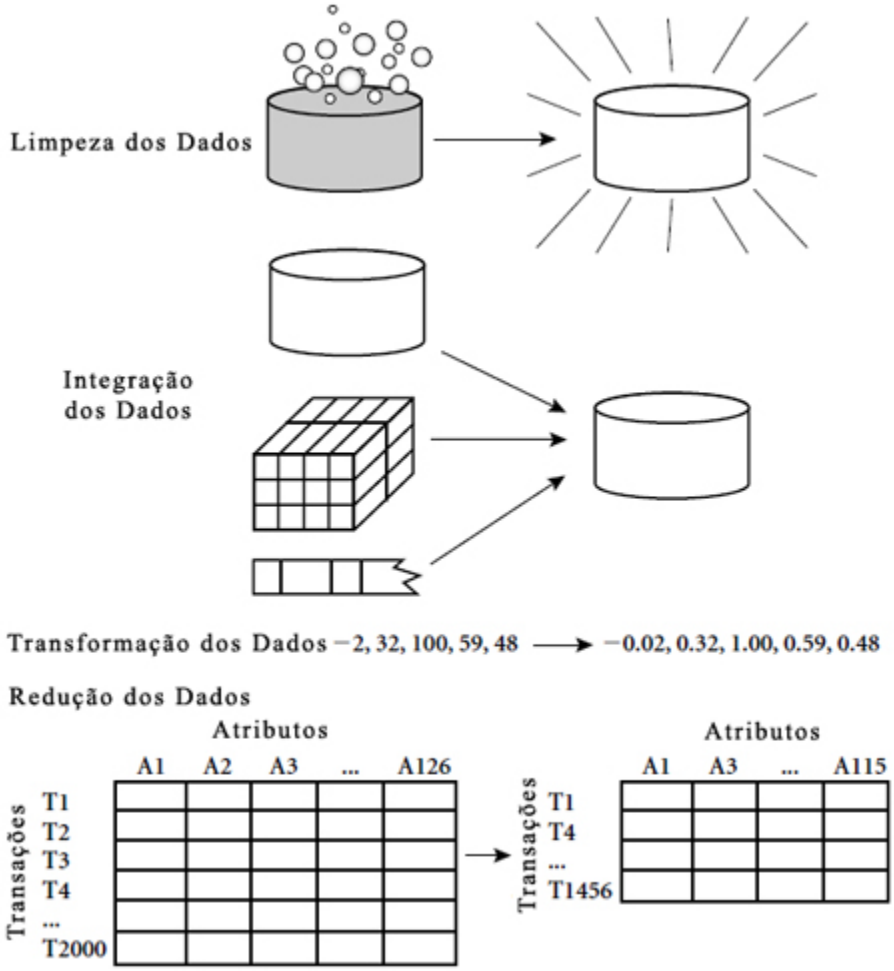
Fonte: SILVA, 2004

A Figura 1 representa todas as fases presentes no processo de KDD. Os passos consistem desde a escolha dos dados a serem analisados, passando pelo seu tratamento e mineração, até que seja possível a extração dos resultados. Porém, para simplificação, esse processo pode ser dividido em três fases principais: Pré-Processamento dos Dados, Mineração de Dados e Pós-Processamento dos Dados.

2.3.1 Pré-Processamento dos Dados

Devido às diversas origens possíveis, é comum que os dados não estejam expressos da maneira correta para que os métodos de Mineração de Dados sejam aplicados diretamente [CAMILO, 2009]. De maneira geral, antes da aplicação dos métodos de Mineração de Dados, ocorrem algumas etapas de preparação das informações, de modo a evitar que valores incorretos possam comprometer o resultado final do estudo. Conhecer os dados que irão ser utilizados para o KDD é essencial para a escolha do método adequado na etapa de mineração de dados. Valores em branco ou nulos, valores viciados e variáveis duplicadas são alguns dos possíveis problemas que podem ser encontrados nos dados. O Pré-processamento consiste em um conjunto de atividades com a finalidade de preparar o banco de dados para o processo de mineração [CAMARGO, 2002], sendo constituído principalmente pelas etapas mostradas na Figura 2.

Figura 2 - Etapas do Pré-Processamento dos Dados



Fonte: SILVA, 2004

Limpeza dos Dados

Frequentemente os dados apresentam discordâncias como registros incompletos, valores errados e dados inconsistentes. Tais discordâncias podem causar alterações no resultado do algoritmo [LANOT, 2012]. Durante essa etapa podem ser utilizadas técnicas de remoção dos valores ou preenchimento com valores padrão [CAMILO, 2009].

Integração dos Dados

Em alguns casos, devido à existência de diversas fontes, surge a necessidade de integração dos dados de maneira que formem um único e consistente repositório. Para isso é necessária uma análise dos dados observando redundâncias, valores conflitantes e dependências [CAMILO, 2009; LANOT, 2012]. Após a análise, os dados são agrupados com base nas similaridades apresentadas, como, por exemplo, um identificador presente em todos os arquivos, e uma nova base é gerada.

Transformação dos Dados

Consiste na utilização de diferentes técnicas com o objetivo de transformar os dados para a forma desejada. Podem ser utilizadas diferentes técnicas de acordo com os objetivos pretendidos. Entre as técnicas existentes, pode-se citar: generalização (converte valores específicos em valores genéricos), normalização (consiste em colocar as variáveis em uma mesma escala), agregação (geração de totalizadores que integram os resultados de diferentes atributos) e criação de novos atributos (gerados a partir de outros atributos existentes) [LANOT, 2012; CAMILO, 2009; SILVA, 2004].

Redução dos Dados

O volume de dados na mineração costuma ser alto, de modo que em alguns casos torna o processo de análise impraticável [CAMILO, 2009]. Dessa forma, características redundantes e dados irrelevantes são eliminados, de maneira a otimizar o tempo de execução do algoritmo [LANOT, 2012].

2.3.2 Mineração de Dados

Segundo CAMARGO(2002), mineração de dados é uma etapa no processo de KDD que consiste na análise de grandes volumes de dados sob diferentes perspectivas, de modo a descobrir novas informações úteis. Esses grandes volumes de dados servem como uma fonte rica para geração de conhecimento através da

utilização de técnicas que envolvem métodos matemáticos, algoritmos e heurísticas para descobrir padrões e regularidades entre os dados estudados.

Apesar de serem utilizados algoritmos para a automatização do processo, ainda é necessária uma análise dos resultados por um humano. Porém, visto que os especialistas podem concentrar seus esforços na análise dos resultados, as tarefas de mineração de dados contribuem de forma significativa no processo de descoberta de conhecimento [CAMILO, 2009].

Dentre os diferentes tipos de técnicas de mineração, pode-se realizar uma classificação entre os utilizados. Os métodos dividem-se em aprendizado supervisionado e não supervisionado. Sua diferença se dá no fato de que os métodos não supervisionados não necessitam de uma pré-categorização onde se defina um alvo, focando a análise na similaridade entre os atributos. Já nos métodos supervisionados, uma variável alvo pré-definida direciona o estudo, de modo que os registros são categorizados em relação a ela [CAMILO, 2009]. Dentre os diferentes tipos de algoritmos de mineração de dados, existem os de Classificação, Clusterização e Regras de associação.

Classificação

De maneira geral, algoritmos de classificação são descritos como supervisionados, tendo seu estudo dirigido por um humano, sendo utilizados para prever os valores de uma variável do tipo categórico [SILVA, 2004]. Essa previsão serve para classificar a qual categoria um registro pertence. Por exemplo, pode-se classificar os clientes de um banco como especiais ou de risco ou um laboratório descobrir quais de seus voluntários podem ser submetidos ao teste de uma nova droga [CAMILO, 2009].

Clusterização

Um *cluster* é uma coleção de registros similares entre si, porém diferente dos registros nos demais agrupamentos. O método de clusterização visa identificar e aproximar os registros similares, não tendo como objetivo classificar, predizer ou estimar o valor de uma variável, apenas identificando os grupos similares [CAMILO, 2009]. Nesse método, ao contrário da classificação, não é necessário que seja

definida uma classe alvo para a tarefa, podendo assim ser classificado como não supervisionado [SILVA, 2004].

Regras de Associação

Consiste em identificar quais atributos possuem relações, representados na forma SE atributo X ENTÃO atributo Y [CAMILO, 2009]. Nesse método, o próprio algoritmo elege os atributos determinantes e os resultantes, gerando as relações entre os atributos [SILVA, 2004]. São amplamente usados em problema do tipo "Cesta de Compras", onde são identificados quais produtos são levados juntos pelos consumidores.

2.3.3 Pós-Processamento dos Dados

Como parte final do processo de KDD, o pós-processamento consiste na etapa de avaliação e interpretação das descobertas. Nessa etapa, as descobertas são selecionadas e ordenadas conforme sua relevância e apresentadas na forma de gráficos ou relatórios para um melhor entendimento [CAMARGO, 2002]. Segundo SILVA(2004), durante essa etapa ocorre a geração de relatórios descrevendo os conhecimentos adquiridos, de forma que possam ser apresentados às partes interessadas.

2.4 Ferramenta Weka[2]

Para realização dos passos de KDD, se faz necessária a utilização de ferramentas que permitam automatização do processo, devido à elevada quantidade de dados envolvidos no processo. Uma dessas ferramentas é o Weka. Segundo [SILVA, 2004], a ferramenta Weka contempla uma série de algoritmos de preparação de dados, de aprendizagem de maquina e validação de resultados. O *software* foi desenvolvido na Universidade de Waikato na Nova Zelândia, sendo escrito em Java e possuindo código aberto. Possui interface gráfica amigável, uma ampla variedade de algoritmos de mineração de dados e seus algoritmos fornecem relatórios com

[2] Disponível em: http://www.cs.waikato.ac.nz/ml/weka/

dados analíticos e estatísticos de acordo com o domínio minerado. A ferramenta Weka foi escolhida para utilização no trabalho por ser desenvolvida na mesma linguagem que a ferramenta proposta, tornando possível uma integração futura.

2.4.1 Indicadores de Qualidade dos Modelos de Classificação

O Weka reúne diversos indicadores que são utilizados em processos de mineração de dados, os quais possibilitam a análise da qualidade dos dados minerados. Com os dados gerados juntamente com a mineração de dados, é possível avaliar se o modelo criado pode ser considerado confiável ou não. Entre os indicadores presentes no Weka, foram utilizados para análise do estudo a Estatística de Kappa e a Matriz de Confusão.

Estatística de Kappa

Para poder realizar uma classificação confiável de algum objeto, é necessário que ele seja avaliado mais de uma vez. E para que seja possível determinar a concordância entre as diferentes classificações de um objeto existe a estatística de Kappa. A estatística de Kappa contabiliza a quantidade de respostas concordantes, determinando se a classificação final não foi obtida ao acaso.

Tabela 1 - Valores de Kappa

Valores de Kappa	Interpretação
<0	Sem concordância
0-0.19	Pouca concordância
0.20-0.39	Concordância razoável
0.40-0.59	Concordância moderada
0.60-0.79	Concordância substancial
0.80-1.00	Concordância quase perfeita

Fonte: BALTAR, 2012

Essa medida assume como valor máximo 1, que significa que ocorreu uma concordância total entre as avaliações, e valores próximo ou até abaixo de 0

representando nenhuma concordância [BALTAR, 2012]. Os diferentes valores de Kappa podem ser interpretados com base na Tabela 1.

Matriz de Confusão

Uma matriz de confusão tem como objetivo realizar uma comparação entre a real classificação de um atributo e a classificação gerada pelo algoritmo utilizado. A matriz é gerada a partir dos diferentes valores assumidos por um atributo, onde é feito um comparativo entre a classificação correta de um atributo e a classificação que foi gerada durante a mineração de dados. Os diferentes valores assumidos pelo atributo são transformados em linhas e colunas. As colunas representam a classificação correta e as linhas representam a classificação gerada pelo algoritmo. Após, a matriz é preenchida com a quantidade de dados classificados. A diagonal principal da matriz exibe a quantidade de dados que foram classificados corretamente, já a quantidade de elementos fora da diagonal representa os erros. Para que ocorra uma classificação sem erros, é necessário que não existam valores fora da diagonal principal da matriz [PESSOA, 2010].

2.4.2 Algoritmo J48

O Weka possui uma grande variedade de algoritmos para mineração de dados. Entre eles, o algoritmo J48, que é a implementação em Java presente no Weka do algoritmo C4.5, permite a criação de modelos de decisão em árvore, o que torna o processo de análise dos resultados mais intuitivo. A construção da árvore ocorre do topo para baixo, onde a seleção de um atributo base irá servir como topo da árvore de decisão. Após a escolha, os dados são divididos em subgrupos, onde os subgrupos são os diferentes valores que o atributo base possui. O processo se repete para cada subgrupo, de maneira que ao final a grande maioria dos atributos pertença a apenas uma classe [MARTINS, 2009].

3 FERRAMENTA DESENVOLVIDA

Neste capítulo será descrito o *software* MineraPampa, o qual realiza as etapas de limpeza, integração, redução e transformação presentes no pré-processamento dos dados. Serão apresentados os métodos e procedimentos adotados para o desenvolvimento do *software*. Adicionalmente, serão descritas as aplicações da ferramenta, as fontes de dados utilizadas e seu desenvolvimento.

3.1 Aplicações da Ferramenta

O MineraPampa tem o objetivo de ser utilizado para tratamento dos dados dos alunos da UNIPAMPA Bagé, podendo também ser utilizado por todas as instituições que fizerem uso do mesmo sistema para armazenamento dos dados dos estudantes, o SIE.

3.2 Ambiente de Desenvolvimento

A ferramenta foi desenvolvida utilizando a linguagem de programação Java, a qual possui as APIs necessárias para tratamento dos dados e por também ser a linguagem de desenvolvimento do Weka, a fim de facilitar uma possível integração futura. A IDE utilizada para programação foi o NetBeans[3], pois possui uma vasta biblioteca de suporte ao programador.

3.3 Análise de Requisitos

Como requisitos funcionais do *software* podem ser descritos:

- O MineraPampa deve realizar o processamento dos dados provenientes do SIE;
- O MineraPampa deve realizar o processamento dos dados provenientes do SiSU;
- O MineraPampa deve realizar a integração dos dados processados do SIE e SiSU;

[3] © Oracle Corporation. Todos os direitos reservados.

- O MineraPampa deve apresentar uma interface intuitiva e de fácil utilização, de modo que um usuário sem conhecimentos técnicos avançados possa opera-lo;
- O MineraPampa deve fornecer opção de escolha do formato de saída dos dados;
- O MineraPampa deve manipular os dados fornecidos pelo usuário e gerar um novo arquivo para utilização posterior em um *software* de mineração de dados.

Como requisitos não funcionais do *software* podem ser descritos:

- O MineraPampa deve ser desenvolvido utilizando ferramentas gratuitas;
- O MineraPampa deve possuir código-fonte aberto.

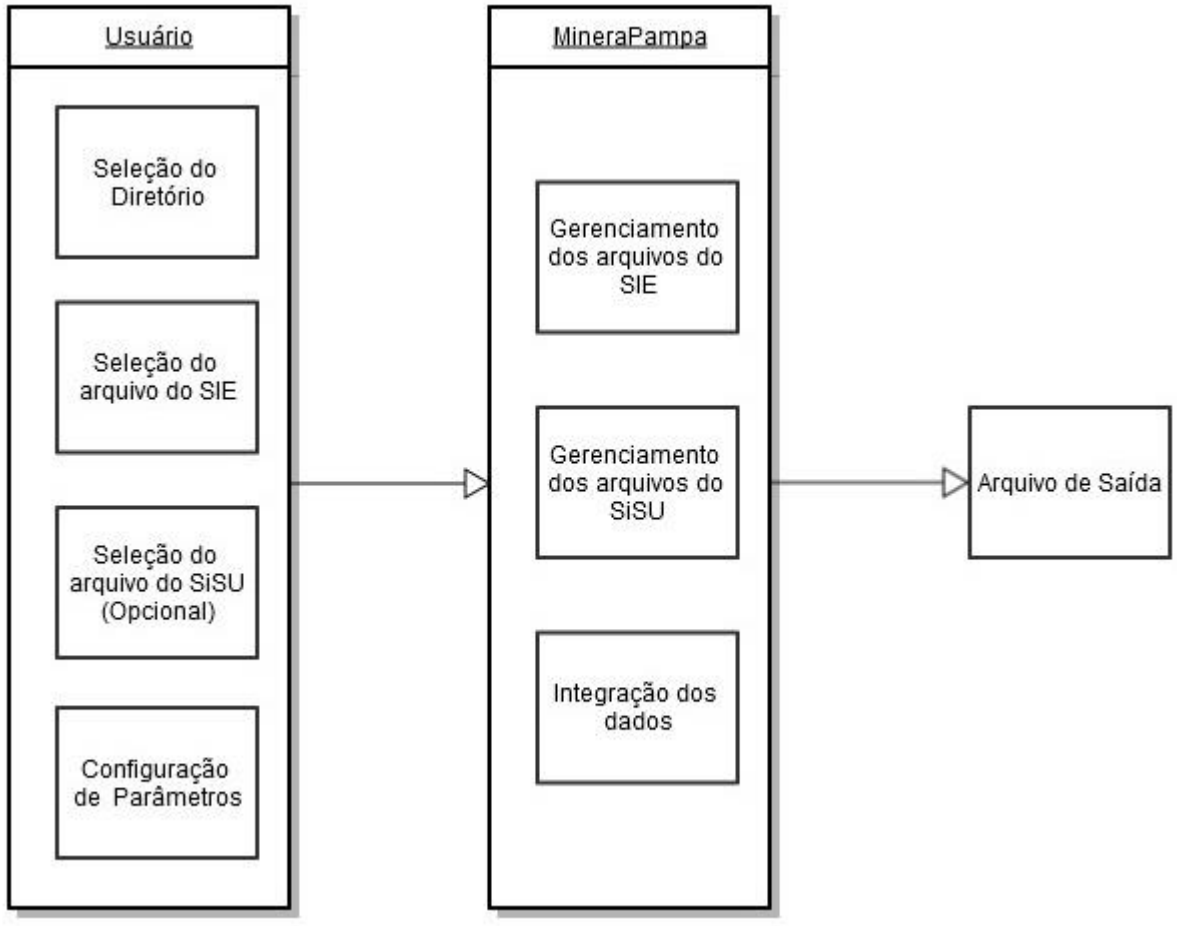
3.4 Processo de Importação de Dados

Os dados foram obtidos dos sistemas SIE e SiSU. O sistema SIE possui todas as informações relativas à vida acadêmica do estudante, como número de matrícula, disciplinas cursadas e notas obtidas. O sistema SiSU contém a nota obtida pelos discentes no ENEM, o qual foi o método de ingresso dos estudantes na universidade à partir do ano de 2010. Para desenvolvimento do trabalho, foram obtidas tabelas contendo informações dos alunos dos cursos de Engenharia de Computação, Engenharia de Alimentos, Licenciatura em Física, Engenharia de Produção, Engenharia Química e Engenharia de Energias Renováveis e Ambiente.

3.5 Diagrama de Arquitetura

O diagrama de arquitetura do *software* desenvolvido é mostrado na Figura 3. Primeiramente, o usuário faz a seleção de um diretório onde estão localizados os arquivos de entrada e onde será armazenado o arquivo de saída. Após, o usuário seleciona quais arquivos deseja processar e quais as configurações que o mesmo deve possuir. O *software* MineraPampa é responsável por realizar o gerenciamento dos arquivos fornecidos com base no parâmetros fornecidos e integrar todos os dados em um único arquivo de saída.

Figura 3 - Diagrama de Arquitetura



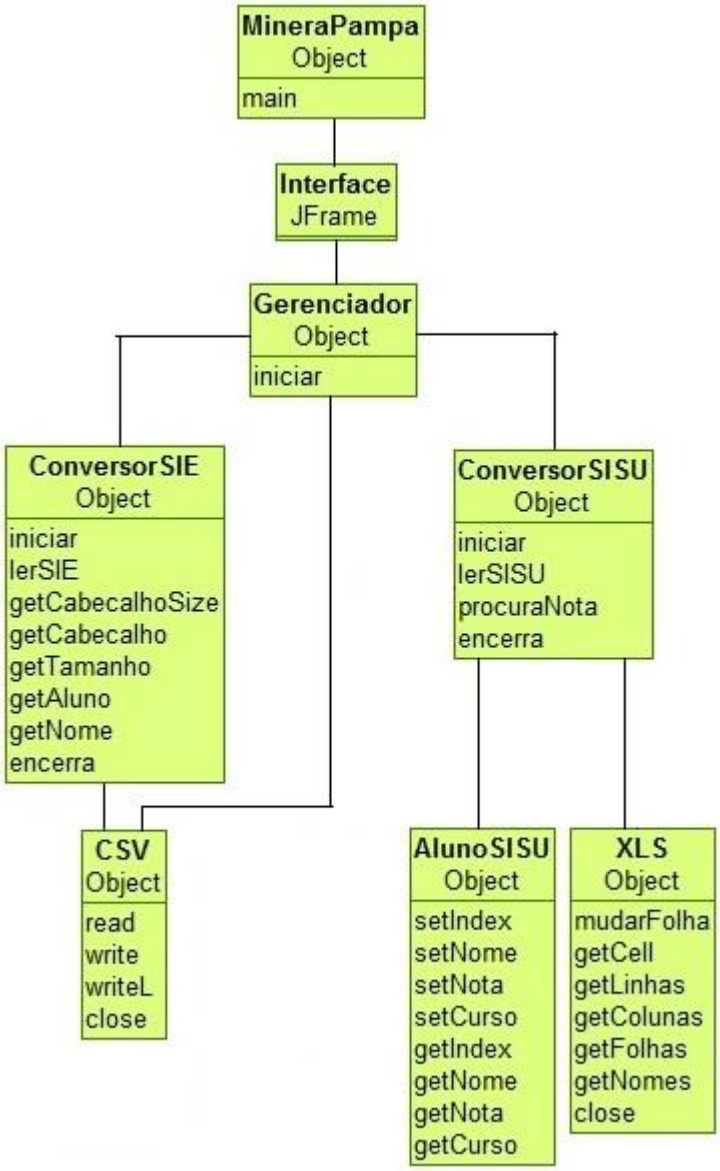
Fonte: Próprio Autor

3.6 Diagrama de Classes

O diagrama de classes do *software* é mostrado na Figura 4. A classe *MineraPampa* apenas realiza a inicialização do *software*. A classe *Interface* realiza a configuração da parte gráfica do *software*, onde ficam localizados os botões para seleção de arquivos a serem processados e as opções disponíveis para o usuário. Após, realiza a instanciação de um objeto da classe *Gerenciador*, o qual irá realizar a integração dos dados que serão manipulados. A classe *ConversorSiSU* é responsável pelo tratamento dos dados provenientes do sistema SiSU. Nela são instanciados dois objetos que serão responsáveis pelo armazenamento das informações processadas (*AlunoSiSU*) e pela realização da leitura dos arquivos contendo os dados (*XLS*). A classe *ConversorSIE* é responsável pelo tratamento dos dados provenientes do sistema SIE. Nela é instanciado um objeto responsável pela

leitura dos dados (*CSV*). Após o processamento dos dados, o objeto instanciado a partir da classe *Gerenciador* realiza a integração dos dados e envia para um objeto instanciado a partir da classe *CSV* para realizar a escrita do arquivo de saída.

Figura 4 - Diagrama de Classes



Fonte: Próprio Autor

3.7 Desenvolvimento

O desenvolvimento pode ser descrito em quatro fases distintas: o gerenciamento dos arquivos do SiSU e SIE, a parte responsável pela escrita no arquivo de saída e a interface que agrupa as partes descritas anteriormente.

3.7.1 Interface

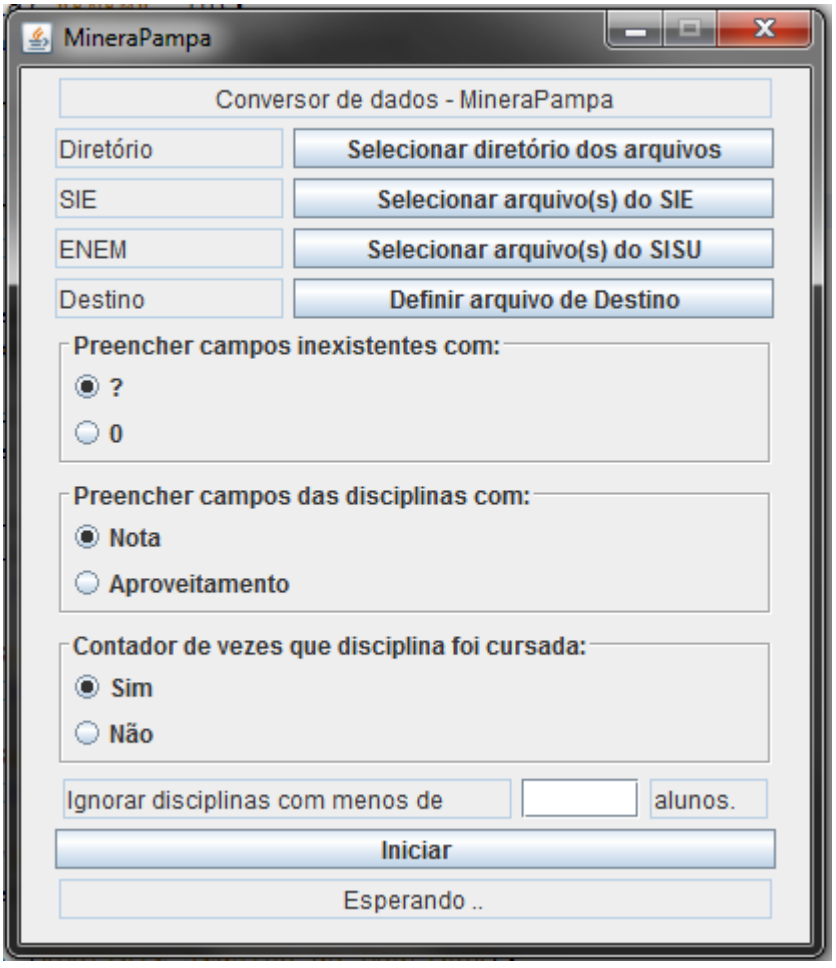
Como parte inicial do desenvolvimento, foi realizada a implementação da interface gráfica do *software*, a qual permite que sejam realizadas as operações de seleção de arquivos desejados e parâmetros para o pré-processamento dos dados.

Primeiramente, foi desenvolvida a janela do MineraPampa, que pode ser verificada na Figura 5, onde ficam localizados todos os botões e opções da ferramenta. Foi utilizada a classe *JFrame* para criação do painel. De modo a facilitar a operação do *software*, foram inseridos botões na interface gráfica de modo a tornar mais intuitiva a operação da ferramenta. Os botões possuem as seguintes funções:

- Selecionar diretório dos arquivos – permite a seleção do diretório que contém os arquivos do SIE e do SiSU, bem como determinar onde o arquivo de saída será salvo;
- Selecionar arquivo(s) do SIE – permite a seleção dos arquivos referentes aos dados do SIE;
- Selecionar arquivos(s) do SiSU – permite a seleção dos arquivos referentes aos dados do SiSU;
- Definir arquivo de destino – permite que o usuário defina o nome do arquivo de saída gerado pelo *software*;
- Preencher campos inexistentes com: “?” ou “0” – opção que permite o usuário definir como serão tratados os campos com dados inexistentes nos arquivos de entrada;
- Preencher campos das disciplinas com: “Nota” ou “Aproveitamento” – permite que o usuário defina qual será a informação contida no campo referente a cada disciplina cursada pelo aluno;

- Contador de vezes que a disciplina foi cursada: Sim ou Não – permite que sejam criadas ou não colunas extras que informam quantas vezes o aluno cursou uma disciplina;
- Ignorar disciplinas com menos de _ alunos – permite que o usuário possa definir um ponto de corte, eliminando disciplinas que foram cursadas por poucos alunos;
- Iniciar – inicia a execução da ferramenta.

Figura 5 - Interface Gráfica do Software



Fonte: Próprio Autor

Para os botões correspondentes à seleção dos arquivos de entrada, estão definidos vetores de arquivos implementados pela classe *File* responsáveis pelo armazenamento dos nomes de arquivos, onde é permitida a escolha de múltiplos arquivos simultaneamente. As opções de seleção de preenchimento do *software* são implementadas com o auxílio da classe *JRadioButton*.

As diferentes opções de preenchimento do arquivo ampliam a quantidade de algoritmos que podem ser utilizados no processo de mineração de dados. Isso ocorre porque os algoritmos operam com diferentes conjuntos de dados. Alguns algoritmos operam com dados categóricos enquanto outros utilizam dados numéricos.

Após ocorrer a definição de todos os arquivos necessários para a execução do pré-processamento, bem como a definição dos parâmetros de preenchimento, é liberada a execução da ferramenta. Nessa etapa, todas as informações necessárias são repassadas para o objeto responsável por esse gerenciamento, o qual é definido pela classe *Gerenciador*. O *Gerenciador* possui objetos responsáveis pelo gerenciamento dos arquivos do SIE, SiSU e de Saída, os quais são descritos na sequência.

3.7.2 Gerenciamento dos arquivos do SiSU

Inicialmente, para tratamento dos dados, foi desenvolvida a parte do *software* responsável pelo gerenciamento do arquivos contendo informações relativas às notas obtidas pelo estudantes no ENEM, os quais são disponibilizados no formato *XLS*. Tais arquivos não apresentam uma estrutura padrão, onde as colunas que contém os atributos variam de um arquivo para o outro. Em alguns arquivos informações como “Inscrição ENEM” estão presentes, em outros existem os campos “Cidade/Estado”, “Data” e “Assinatura”. Outra característica presente nos arquivos é a inconsistência no preenchimento dos campos, pois mesmo com campos como “Cidade/Estado” estando presentes, não existe nenhuma informação ali disponível. A relação de todos os campos presentes nos arquivos são apresentados na Tabela 2 abaixo:

- Os campos “CPF” e “Inscrição ENEM” não devem ser utilizados devido à necessidade de sigilo das informações do aluno;
- Os campos “Data” e “Assinatura” não são informações relevantes no processo de mineração de dados;
- O campo “Cidade/Estado” não está preenchido;
- O campo “Classificação” está relacionado diretamente à nota obtida no ENEM, sendo necessária apenas a inclusão da nota.

Tabela 2 - Dados dos Arquivos do SiSU

Campo	Descrição
Classificação	Classificação do aluno no SiSU.
CPF	CPF do aluno.
Nome	Nome do aluno.
Inscrição ENEM	Número de inscrição no ENEM.
Nota	Nota obtida no ENEM.
Cidade/Estado	Cidade e Estado do aluno.
Data	Data de realização da matricula.
Assinatura	Assinatura do aluno.

Fonte: Próprio Autor

Devido à tais particularidades, as únicas informações relevantes que podem ser extraídas desses arquivos são “Nome” e “Nota”, onde o nome é necessário para futuras comparações quando a nota for atribuída ao aluno correspondente.

Para execução desses passos, primeiramente o objeto Gerenciador instancia o objeto *GerenciadorSIE* enviando os arquivos escolhidos pelo usuário e o caractere que deve ser utilizado para preenchimento das informações inexistentes, “?” ou “0”. O objeto *GerenciadorSIE* realiza a abertura de um arquivo por iteração, onde após o termino da iteração fecha o arquivo atual e realiza os mesmos procedimentos para o próximo arquivo até que todos sejam processados.

Quando um arquivo é aberto, o mesmo contém os dados dos alunos separados por curso, onde cada curso é representado por uma folha dentro do arquivo. Sendo assim, além de ser necessária realizar a abertura de cada arquivo de forma separada, também deve-se abrir as várias folhas presentes em cada arquivo. Ambos processos ocorrem através do controle do número de arquivos e do número de folhas em cada arquivo por variáveis que funcionam como contadores.

Após a abertura do arquivo e de uma de suas folhas, as colunas contendo atributos são percorridas, de modo a comparar a informação contida no arquivo com os atributos desejados, que são “Nome” e “Nota”. Essa comparação é necessária para definir os índices onde estão presentes as informações, pois tais índices variam entre os arquivos, sendo necessária a realização desse passo a cada novo arquivo aberto pela ferramenta. Após, são percorridas as linhas, onde as informações de cada aluno são adicionadas em uma lista do tipo *ArrayList*. Quando a folha é

analisada até seu final, ocorre a abertura da próxima folha presente no arquivo. A etapa de processamento dos dados do SiSU acaba quando todas as folhas de todos os arquivos forem analisadas. Ao final desse processo, a lista contém todos os nomes dos alunos com suas respectivas notas.

A lista fica disponível para futuras consultas, onde é necessário enviar o nome do aluno que se deseja obter a nota. O objeto faz uma varredura pela lista comparando os nomes, onde caso seja encontrado o aluno a nota é retornada, ou caso contrário, o caractere destinado a preenchimento de informações inexistentes. Após a etapa de processamento dos arquivos do SiSU, é iniciado o processamento dos dados do SIE.

3.7.3 Gerenciamento dos arquivos do SIE

Nessa etapa ocorre o mais importante dos processamentos, pois nela a estrutura do arquivo é totalmente alterada. Além de serem realizados alguns passos mais simples, como limpeza, redução e integração dos dados, ocorre uma transformação na estrutura, onde dados que eram tuplas passam a assumir o papel de atributos. Isso permite que todos os dados de um aluno possam ser descritos em apenas uma linha do arquivo final, ao contrário do arquivo origem, quando um mesmo aluno possuía várias linhas com dados redundantes, onde as informações relativas à disciplina eram as únicas a possuírem uma alteração.

Na Tabela 3 localizada abaixo são descritos os atributos presentes nos arquivos provenientes do SIE:

- “ID_PESSOA”, “NOME_PESSOA”, “ID_ALUNO”, “MATR_ALUNO” são todas referências ao aluno, sendo necessária manter apenas uma delas ao final do processo, onde “MATR_ALUNO” foi a opção escolhida, pois mantém o nome do aluno em sigilo;
- “NUM_VERSAO” refere-se ao ano em que o cadastro foi alterado pela ultima vez, não sendo necessária sua inclusão;
- “NOME_CURSO”, “COD_CURSO”, “ID_VERSAO_CURSO”, “ID_CURSO_ALUNO” são todas referências ao curso, sendo necessária manter apenas uma delas ao final do processo, onde “NOME_CURSO” foi a opção escolhida, pois contém o nome completo do curso;

Tabela 3 - Dados dos Arquivos do SIE

Campo	Descrição
ID_PESSOA	Número de identificação.
NOME_PESSOA	Nome do aluno.
ID_ALUNO	Número de identificação.
MATR_ALUNO	Número de matrícula.
NUM_VERSAO	Ano de versão do cadastro.
NOME_CURSO	Nome do curso.
COD_CURSO	Código do curso.
ID_VERSAO_CURSO	Número de identificação do curso.
ANO	Ano em que a disciplina foi cursada.
COD_ATIV_CURRIC	Código da disciplina.
NOME_ATIV_CURRIC	Nome da disciplina.
CREDITOS	Número de créditos da disciplina.
MEDIA_FINAL	Média final na disciplina.
DESCR_SITUACAO	Descrição da situação do aluno.
PERIODO	Período que a disciplina foi cursada.
ID_CURSO_ALUNO	Número de identificação do curso.
SITUACAO_ITEM	Descrição da situação na disciplina.
CH_TEORICA	Carga horária teórica da disciplina.
CH_PRATICA	Carga horária prática da disciplina.
TOTAL_CARGA_HORARIA	Total de carga horária da disciplina.
FORMA_INGRESSO	Forma de ingresso.
ANO_INGRESSO	Ano de ingresso.
FORMA_EVASÃO	Forma de evasão.
ANO_EVASÃO	Ano de evasão
SEXO	Sexo do aluno.

Fonte: Próprio Autor

- “ANO” e “PERIODO” referem-se ao ano e semestre em que a disciplina foi cursada. Apesar de serem informações relevantes, seria necessária a criação de dois novos atributos extras ao lado de cada atributo referente a uma disciplina, onde o numero de colunas presentes no arquivo final seria muito alto. Portanto, foi decidido não incluir esses campos no arquivo final;

- “COD_ATIV_CURRIC”, “NOME_ATIV_CURRIC”, “CREDITOS”, “CH_TEORICA”, “CH_PRATICA” e “TOTAL_CARGA_HORARIA” são todas referências à disciplina, sendo necessária manter apenas uma delas ao final do processo, onde “NOME_ATIV_CURRIC” foi a opção escolhida, pois contém o nome completo da disciplina. Nesse ponto ocorre a grande transformação dos arquivos relacionados ao SIE, pois essa informação irá tornar-se um atributo no arquivo final;
- “MEDIA_FINAL” e “DESCR_SITUACAO” fazem referência ao desempenho do aluno na disciplina em questão. Apenas uma dessas informações será colocada no arquivo final, onde a opção marcada na interface gráfica do *software* será analisada;
- “FORMA_INGRESSO”, “ANO_INGRESSO”, “FORMA_EVASAO” e “SEXO” são informações relevantes, portanto estão presentes no arquivo final;
- “ANO_EVASAO” estava incluso no arquivo final durante a primeira parte do trabalho, porém foi substituído por “ANOS_CURSADOS” que terá seu cálculo explicado posteriormente.

O processo inicia após a abertura do arquivo contendo as informações do SIE, o qual possui o formato *CSV*. Inicialmente, é realizada uma varredura por todo arquivo, analisando os valores do atributo “NOME_ATIV_CURRIC”, pois essas tuplas serão necessárias para formação de novos atributos no arquivo de saída. Esse processo começa pela leitura de cada linha presente no arquivo, onde a mesma é quebrada em uma *String* com o comando s*plit*, o qual permite a separação das informações onde ocorre a presença de um caractere previamente definido. O caractere em questão é “,”, caractere padrão para separação de campos em um arquivo *CSV*. Após, é realizada a comparação do nome da disciplina com todos os nomes presentes em uma lista de disciplinas. Como essa lista não possui nenhum valor durante a primeira execução, o primeiro valor sempre é adicionado, e os valores subsequentes são analisados. Sempre que é realizada a leitura de uma disciplina previamente presente na lista, a mesma é descartada. Quando ocorre a leitura de uma disciplina que ainda não está presente na lista, a mesma é adicionada. A comparação ocorre através da varredura da lista, onde são realizadas comparações entre a disciplina lida e a disciplina contida na lista.

Durante o mesmo laço de execução, uma segunda lista é preenchida com a quantidade de alunos que cursaram a disciplina. Quando uma nova disciplina é encontrada, ela é adicionada na lista como uma informação nova. O mesmo acontece na lista de quantidades, onde é inserido o número “1” na lista de quantidades, onde o índice para acesso na lista de quantidades é o mesmo para a lista de nomes, garantindo que um mesmo índice possua informações sobre a mesma disciplina nas duas listas. Quando uma disciplina é encontrada na lista de nomes, ocorre o incremento do número contido na lista de quantidades.

Ao final dessa etapa, são obtidas duas listas de todas as disciplinas presentes no arquivo do SIE e a quantidade de alunos que cursou cada uma delas. A lista contendo os nomes de disciplinas é, juntamente com as informações escolhidas previamente, os novos atributos do arquivo de saída. A maneira como é realizada a passagem desses valores para o arquivo final será descrita na próxima subseção desse capítulo.

O próximo passo consiste em reunir todas as informações de um único aluno em um vetor. Para isso, o ponteiro de leitura é reposicionado no inicio e as linhas começam a serem lidas novamente. Porém, dessa vez a tupla analisada é a de nome do aluno, onde enquanto o nome for igual, as informações são coletadas. Durante esse processo, um vetor do tipo *String* é totalmente preenchido com o caractere para dados inexistentes. Conforme as informações são obtidas, elas sobrepõem esse caractere especial. Nos atributos onde não existem informações correspondentes, os campos já estão preenchidos com o valor final.

Inicialmente, é lida uma linha e o nome do aluno é atribuído à uma variável temporária. Todo o restante de suas informações também é lido nesse passo, como curso, forma de ingresso e sexo. Nesse passo também é feita a leitura da nota do aluno ou seu aproveitamento na disciplina contida nessa linha. Para determinar a posição do vetor que deve ser preenchida com o dado da disciplina, seu nome é enviado para uma função que compara o nome da disciplina atual com a correspondente na lista armazenada previamente em memoria. Quando é encontrada, seu índice é retornado e o dado da disciplina é colocado na posição do vetor correspondente à coluna no arquivo final. Durante a leitura das próximas linhas do mesmo aluno, apenas é necessário obter a disciplina cursada para determinar a posição do vetor que a informação de nota ou aproveitamento será posta. Durante a mesma etapa, ocorre o preenchimento de um segundo vetor, o qual é responsável

por armazenar as informações relativas à quantidade de vezes que o aluno cursou cada disciplina. Inicialmente o vetor é preenchido com zeros e, conforme o preenchimento do vetor contendo as notas ocorre, o vetor de quantidades é incrementado.

Após o final dos dados de um aluno se esgotarem, ocorre a pesquisa pela nota do ENEM. Para isso, o nome do aluno é enviado para uma função do objeto *GerenciadorSiSU* que retorna a nota obtida pelo aluno ou o caractere para preenchimento de informações inexistentes. Nos campos relativos à notas obtidas, o caractere separador de casas decimais é “,”, o qual é o mesmo caractere utilizado para separação de campos no arquivo final. Portanto, foi necessário realizar a substituição da vírgula por um ponto “.” com auxilio do comando *replace*. Esses passos são repetidos para todos alunos, até o termino do processamento do arquivo. O próximo passo é a escrita dos dados no arquivo de saída.

3.7.3.1 Novos dados e tratamento de casos especiais

Além do tratamento realizado na estrutura do arquivo, foi necessária a criação de mecanismos para manipular dados que possuam valores errôneos e/ou desnecessários. Também foram criadas soluções para a geração de novos dados com base nos disponíveis nas tabelas obtidas do SIE.

Durante a primeira parte do trabalho, foi observado que o campo “ANO_EVASAO” não contribuiu com informações significativas durante o processo de mineração de dados. Esse campo foi substituído por “ANOS_CURSADOS”, o qual possui a quantidade de anos que o aluno está matriculado em seu curso na UNIPAMPA. Com base em “ANO_INGRESSO” e a data atual do sistema operacional que está rodando o *software*, foi realizado o calculo desse novo campo e o mesmo foi utilizado no arquivo final.

Outro fator que foi considerado necessário durante a primeira parte do trabalho foi a criação de um campo contendo a quantidade de alunos que cursaram determinada disciplina. Como descrito anteriormente, esse campo é calculado pelo *software* e sua inclusão no arquivo de saída é opcional. Na interface do programa é possível determinar se deve ser realizada a inclusão ou não dessa informação, pois sua inclusão acarreta na duplicação da quantidade de informações relativas à

disciplinas no arquivo final, pois é gerada uma nova coluna para cada disciplina presente.

Como tratamento de casos especiais, foram criados mecanismos para tratar problemas que foram percebidos durante o TCC I. Quando um aluno cursa alguma disciplina mais de uma vez, o arquivo do SIE apresenta uma linha para cada vez que foi cursada. Dessa maneira, um mesmo aluno possuía diversas notas para uma mesma disciplina. Foi criado um teste para inserir no arquivo final apenas a nota da aprovação do aluno, desconsiderando casos de reprovação ou trancamento.

Outra peculiaridade é que o arquivo do SIE apresenta os valores “0” e “100000” para representar os seguintes casos:

- Reprovado por frequência;
- Trancamento;
- Aproveitamento;
- Atividades Complementares de Graduação.

Os valores apresentados interferiam no resultado da mineração de dados. Com isso, foram desconsiderados no arquivo final.

Como ultimo ponto a ser destacado, alguns alunos mudam de curso ou cursam disciplinas de outros cursos como horas complementares durante a graduação. Com isso, algumas disciplinas presentes no arquivo de saída possuíam poucos alunos com notas atribuídas. Foi criado, segundo um valor determinado pelo usuário, um mecanismo para realizar a eliminação de disciplinas que foram cursadas por menos alunos que o valor informado. Para isso, durante a etapa de preenchimento das disciplinas que serão utilizadas é utilizada lista contendo a quantidade de alunos que cursou a cadeira.

3.7.4 Gerenciamento do arquivo de Saída

Como parte final da execução da ferramenta, ocorre a escrita no arquivo de saída em *CSV*. Essa escrita ocorre em dois momentos distintos: quando a etapa de pesquisa de disciplinas é concluída e quando todas as informações de um aluno são coletadas. Para ambos os casos o procedimento é igual. É fornecida a um objeto da classe *CSV* uma *String*, onde os vetores contendo os atributos e os de cada aluno são concatenados em uma *String*, separados por “,” e então escritos no arquivo final. O formato do arquivo final pode ser visualizado na Tabela 4.

Após o termino da escrita, todos os arquivos utilizados são fechados e a etapa de pré-processamento dos dados está concluída. O próximo passo consiste na mineração de dados, que é descrita no próximo capitulo.

Tabela 4 - Dados dos Arquivos de Saída

Campo	**Descrição**
MATRICULA	Numero de matricula.
CURSO	Nome do curso.
FORMA_INGRESSO	Forma de ingresso.
ANO_INGRESSO	Ano de ingresso.
FORMA_EVASAO	Forma de evasão.
ANOS_CURSADOS	Quantidade de anos que o aluno cursou.
SEXO	Sexo do aluno.
DISCIPLINAS	Relação de todas as disciplinas.
DISCIPLINAS_QT	Quantidade de vezes que a disciplina foi cursada.
NOTA_ENEM	Nota obtida no ENEM.

Fonte: Próprio Autor

4 MINERAÇÃO DE DADOS

A etapa de mineração de dados consiste na aplicação, através da ferramenta Weka, de algoritmos nos dados processados anteriormente pelo software desenvolvido, de modo que possam ser detectados padrões entre os estudantes evadidos. Como o estudo é dirigido para compreensão da evasão, foram realizados experimentos somente com algoritmos do tipo de classificação, pois é possível definir um foco de estudo através de parâmetros de configuração dos algoritmos. Os experimentos foram realizados utilizando oito combinações de pré-processamento para cada arquivo de informações. As combinações listadas abaixo são possíveis com base nas diferentes combinações de configuração da ferramenta desenvolvida:

- Preenchimento com “0” e “Notas”;
- Preenchimento com “0” e “Aproveitamento”;
- Preenchimento com “?” e “Notas”;
- Preenchimento com “?” e “Aproveitamento”.

As quatro combinações listadas acima juntamente com a opção de inserir ou não colunas indicando a quantidade de vezes que o aluno cursou cada disciplina possibilitaram a geração dos oito arquivos utilizados para a mineração de dados. O ponto de corte para número de disciplinas foi definido como “5” utilizando como base os experimentos realizados na primeira parte do trabalho, onde a grande maioria das disciplinas pertencentes a outros cursos tinham sido cursadas menos de cinco vezes.

Os cursos analisados estão separados nos subcapítulos apresentados abaixo. Os algoritmos utilizados podem apresentar mais de um resultado quando aplicados em arquivos diferentes. Os diferentes resultados para um mesmo algoritmo estão listados em experimentos com diferentes numerações, acompanhados da configuração de arquivo utilizada para o procedimento.

Todos os arquivos gerados foram submetidos a aplicações dos algoritmos de mineração de dados. Durante a mineração de dados dos arquivos relativos ao curso de Engenharia de Computação foram utilizados todos os algoritmos disponíveis no Weka. Para a realização dos experimentos seguintes, foram utilizados os algoritmos que apresentaram os melhores resultados durante a mineração de dados realizada previamente. Com isso, justifica-se a predominância dos algoritmos

FilteredClassifier, *JRip*, *PART* e *J48* nos experimentos, pois geraram os melhores resultados para todos os cursos analisados.

Adicionalmente, cada experimento apresenta os dados que indicam a qualidade da mineração de dados, onde é possível determinar se a regra encontrada possui um grau de confiabilidade relevante.

4.1 Engenharia de Computação

Os resultados obtidos durante a mineração de dados relativa aos estudantes de Engenharia de Computação estão descritos nos experimentos abaixo.

Quadro 1 – EC – Algoritmo FilteredClassifier – Experimento 1

```
ANOS_CURSADOS = '(-inf-0.5]'
|   ALGORITMOS E PROGRAMAÇÃO = '(-inf-3.05]': Abandono (61.6/8.0)

ANOS_CURSADOS = '(0.5-3.5]'
|   ANO_INGRESSO = '(-inf-2011.5]'
|   |   FORMA_INGRESSO = Portador de Diploma: Abandono (7.0)
|   |   FORMA_INGRESSO = Transferência: Abandono (11.0/2.0)

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances         342               74.6725 %
Incorrectly Classified Instances       116               25.3275 %
Kappa statistic                          0.5693

=== Confusion Matrix ===
   a   b   c   d   e   f   g   h   i   <-- classified as
 188  20   0   1   1   0   0   0   0 |   a = Aluno Regular
  20 150   4   1   0   0   0   0   0 |   b = Abandono
   2  22   3   0   0   0   0   0   0 |   c = Cancelamento
   8   7   0   1   0   0   0   0   0 |   d = Transf. Interna Por Reopção de Curso
  11   0   0   0   0   0   0   0   0 |   e = Formado
   1   6   1   0   0   0   0   0   0 |   f = Transferência
   0   5   0   0   0   0   0   0   0 |   g = Desligamento
   0   5   0   0   0   0   0   0   0 |   h = Transferência Interna
   1   0   0   0   0   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 1 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Os resultados estão divididos em duas partes, relacionadas à quantidade de anos cursados. Entre os alunos que cursaram até 0,5 anos e obtiveram notas entre 0 e 3,05 em Algoritmos e Programação, ocorreram 52 abandonos dentre os 61 estudantes que pertencem a regra. Com relação aos alunos que cursaram entre 0,5 e 3,5 anos, realizando o ingresso na faculdade até 2011, é demonstrada uma ligação entre abandono e a forma de ingresso. Alunos que ingressaram na universidade como Portadores de Diploma ou Transferência acabaram evadindo, exceto em dois casos.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que as classificações realizadas como abandono foram correta na grande maioria dos casos, pois apenas 25 estudantes que abandonaram a faculdade não foram classificados corretamente.

Quadro 2 – EC – Algoritmo FilteredClassifier – Experimento 2

```
ANO_INGRESSO = '(-inf-2011.5]'
|   DESENHO TÉCNICO I-QT = '(-inf-0.5]'
|   |   SISTEMAS OPERACIONAIS-QT = '(-inf-0.5]': Abandono (203.0/56.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         354               77.2926 %
Incorrectly Classified Instances       104               22.7074 %
Kappa statistic                          0.6229

=== Confusion Matrix ===

   a    b   c   d   e   f   g   h   i   <-- classified as
 183   21   0   1   5   0   0   0   0 |   a = Aluno Regular
   7  162   0   4   0   2   0   0   0 |   b = Abandono
   3   22   0   0   0   2   0   0   0 |   c = Cancelamento
   4    9   0   3   0   0   0   0   0 |   d = Transf. Interna Por Reopção de Curso
   5    0   0   0   6   0   0   0   0 |   e = Formado
   3    4   0   1   0   0   0   0   0 |   f = Transferência
   0    5   0   0   0   0   0   0   0 |   g = Desligamento
   0    5   0   0   0   0   0   0   0 |   h = Transferência Interna
   1    0   0   0   0   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 2 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Nesse experimento é possível analisar que a evasão ocorre durante os primeiros semestres do curso. O estudo demonstra que entre os alunos que ingressaram até o ano de 2011 na faculdade e não cursaram as disciplinas de Desenho Técnico I e Sistemas Operacionais, que estão dispostas a partir do 4° semestre na grade curricular do curso, ocorreram 147 casos de abandono dentre os 203 estudantes que se encaixam na regra.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que 162 estudantes foram classificados corretamente e apenas 13 estudantes que abandonaram a faculdade apresentaram classificação errônea.

Quadro 3 – EC – Algoritmo JRip – Experimento 1

```
(INTRODUÇÃO A ARQUITETURA DE COMPUTADORES <= 3.5) and (ANOS_CURSADOS <= 1)
=> FORMA_EVASÃO=Abandono (102.0/14.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         374               81.6594 %
Incorrectly Classified Instances       84                18.3406 %
Kappa statistic                        0.692

=== Confusion Matrix ===

  a   b   c  d  e  f  g  h  i   <-- classified as
204   2   1  1  1  1  0  0  0 |  a = Aluno Regular
 13 158   2  2  0  0  0  0  0 |  b = Abandono
  4  22   1  0  0  0  0  0  0 |  c = Cancelamento
  7   8   0  1  0  0  0  0  0 |  d = Transf. Interna Por Reopção de Curso
  1   0   0  0 10  0  0  0  0 |  e = Formado
  0   7   1  0  0  0  0  0  0 |  f = Transferência
  0   5   0  0  0  0  0  0  0 |  g = Desligamento
  0   5   0  0  0  0  0  0  0 |  h = Transferência Interna
  1   0   0  0  0  0  0  0  0 |  i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 3 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Neste experimento é demonstrada uma relação entre o desempenho na disciplina de Introdução a Arquitetura de Computadores e o abandono do curso.

Entre os alunos que obtiveram notas iguais ou inferiores a 3,5 na disciplina e cursaram até um ano da faculdade, ocorreram 88 casos de abandono da universidade.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que apenas 17 estudantes que abandonaram a faculdade não foram classificados corretamente.

Quadro 4 – EC – Algoritmo JRip – Experimento 2

```
(LABORATÓRIO DE FÍSICA I-QT >= 1) and (INTRODUÇÃO A ARQUITETURA DE
COMPUTADORES-QT >= 2) => FORMA_EVASÃO=Transf. Interna Por Reopção de Curso
(5.0/1.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         372               81.2227 %
Incorrectly Classified Instances        86               18.7773 %
Kappa statistic                          0.6887

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 202   2   1   2   1   2   0   0   0 |   a = Aluno Regular
  13 155   5   2   0   0   0   0   0 |   b = Abandono
   6  20   0   1   0   0   0   0   0 |   c = Cancelamento
   5   7   0   4   0   0   0   0   0 |   d = Transf. Interna Por Reopcão de Curso
   0   0   0   0  11   0   0   0   0 |   e = Formado
   1   6   1   0   0   0   0   0   0 |   f = Transferência
   0   5   0   0   0   0   0   0   0 |   g = Desligamento
   1   3   0   1   0   0   0   0   0 |   h = Transferência Interna
   1   0   0   0   0   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 4 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Aproveitamento” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O experimento demonstra um provável perfil existente entre os alunos que realizam a reopção de curso dentro da universidade. No Quadro 4 está expressa uma relação que ocorreram 4 casos de mudança de curso entre os alunos que cursaram uma ou mais vezes Laboratório de Física I e pelo menos duas vezes a disciplina de Introdução a Arquitetura de Computadores. Essa estatística demonstra

que reprovações em Introdução a Arquitetura de Computadores pode ser uma característica comum entre os estudantes que mudam de curso.

A estatística de Kappa demonstra que a concordância dos resultados é substancial, porém a matriz de confusão mostra que um número elevado de estudantes que realizaram a reopção de curso estão classificados incorretamente, onde apenas 4 entre 16 foram analisados de maneira correta.

Quadro 5 – EC – Algoritmo JRip – Experimento 3

```
(FÍSICA I >= 6.75) and (ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES I <= 5.1) and
(TÉCNICAS DIGITAIS <= 2.3) => FORMA_EVASÃO=Transf. Interna Por Reopção de Curso
(12.0/3.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         381               83.1878 %
Incorrectly Classified Instances        77               16.8122 %
Kappa statistic                          0.7204

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 205   0   0   3   1   1   0   0   0 |   a = Aluno Regular
  12 160   0   2   0   1   0   0   0 |   b = Abandono
   5  22   0   0   0   0   0   0   0 |   c = Cancelamento
   2   7   1   6   0   0   0   0   0 |   d = Transf. Interna Por Reopção de Curso
   1   0   0   0  10   0   0   0   0 |   e = Formado
   2   4   1   1   0   0   0   0   0 |   f = Transferência
   0   5   0   0   0   0   0   0   0 |   g = Desligamento
   0   4   0   1   0   0   0   0   0 |   h = Transferência Interna
   1   0   0   0   0   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 5 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Neste experimento é possível analisar mais um provável perfil existente entre os alunos que realizam a reopção de curso. Notas inferiores a 5,1 em Arquitetura e Organização de Computadores I e 2,3 em Técnicas Digitais, aliadas a um desempenho igual ou superior a 6,75 em Física I são características comuns entre 9 estudantes que alteraram seu curso na universidade. Isso demonstra uma relação entre reprovações em disciplinas especificas do curso de Engenharia de Computação e a desistência de cursar o mesmo.

A estatística de Kappa demonstra que a concordância dos resultados é substancial, porém a matriz de confusão mostra que um número razoável de estudantes que realizaram a reopção de curso estão classificados incorretamente, onde 6 entre 16 foram analisados de maneira correta.

Quadro 6 – EC – Algoritmo JRip – Experimento 4

```
(ANOS_CURSADOS <= 1) and (ANO_INGRESSO <= 2012) and (INTRODUÇÃO A
ENGENHARIA DE COMPUTAÇÃO <= 4.55) => FORMA_EVASÃO=Abandono (41.0/3.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         371               81.0044 %
Incorrectly Classified Instances        87               18.9956 %
Kappa statistic                          0.6823

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 205   0   0   3   2   0   0   0   0 |   a = Aluno Regular
  19 152   1   2   0   1   0   0   0 |   b = Abandono
   9  17   0   1   0   0   0   0   0 |   c = Cancelamento
   5   6   1   4   0   0   0   0   0 |   d = Transf. Interna Por Reopção de Curso
   1   0   0   0  10   0   0   0   0 |   e = Formado
   1   6   1   0   0   0   0   0   0 |   f = Transferência
   0   5   0   0   0   0   0   0   0 |   g = Desligamento
   2   2   0   1   0   0   0   0   0 |   h = Transferência Interna
   1   0   0   0   0   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 6 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Notas” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

No Quadro 6 é demonstrada uma relação presente entre reprovação em uma disciplina especifica e o abandono. Ocorreram 38 casos de abandono entre os estudantes que cursaram até um ano de faculdade, ingressaram até o ano de 2012 e obtiveram desempenho igual ou inferior a 4,55 em Introdução a Engenharia de Computação, uma das bases do curso.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 7 – EC – Algoritmo PART – Experimento 1

```
ANOS_CURSADOS <= 0 AND
ALGORITMOS E PROGRAMAÇÃO <= 4.35: Abandono (62.65/8.0)

ANOS_CURSADOS <= 2 AND
SEXO = M AND
FÍSICA I <= 3.6: Abandono (70.8/10.72)

ANOS_CURSADOS <= 3 AND
CALCULO I <= 4.6 AND
SEXO = M: Abandono (22.84/1.28)

ANOS_CURSADOS > 3 AND
DESENHO TÉCNICO I <= 6.5: Abandono (10.33/2.67)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         385               84.0611 %
Incorrectly Classified Instances        73               15.9389 %
Kappa statistic                          0.7375

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 210   0   0   0   0   0   0   0   0 |   a = Aluno Regular
   0 163   3   3   5   0   1   0   0 |   b = Abandono
   0  24   3   0   0   0   0   0   0 |   c = Cancelamento
   0  12   0   3   1   0   0   0   0 |   d = Transf. Interna Por Reopção de Curso
   0   1   0   4   6   0   0   0   0 |   e = Formado
   0   6   1   1   0   0   0   0   0 |   f = Transferência
   0   5   0   0   0   0   0   0   0 |   g = Desligamento
   0   5   0   0   0   0   0   0   0 |   h = Transferência Interna
   0   0   0   0   1   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 7 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Este experimento mostra que a evasão no curso de Engenharia de Computação, apesar de ter forte ligação com o primeiro ano de curso, apresenta casos onde alunos desistiram depois de anos de estudos. Na primeira parte do experimento é possível analisar que alunos que obtiveram notas inferiores a 4,35 em algoritmos tiveram uma alta taxa de abandono, onde dos 62 alunos que se enquadram nesse caso, apenas 8 não abandonaram o curso. Na segunda parte existe a relação com Física I, onde alunos do sexo masculino e que cursaram até dois anos de faculdade, aliados a notas inferiores a 3,6 em Física I, acabaram

abandonando a universidade na maioria dos casos. Na terceira parte ocorre um caso similar, porém onde alunos do sexo masculino que cursaram até três anos e obtiveram nota inferior a 4,6 em Calculo I evadiram. Por ultimo, ocorre a demonstração que alunos com mais de três anos de faculdade e notas inferiores a 6,5 em Desenho Técnico I evadiram em quase todos os casos.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 8 – EC – Algoritmo PART – Experimento 2

```
CALCULO I <= 4.6 AND
LABORATÓRIO DE PROGRAMAÇÃO I <= 2.8: Abandono (14.0/3.0)

INTRODUÇÃO A ENGENHARIA DE COMPUTAÇÃO <= 3.5 AND
FÍSICA I > 6.33: Transf. Interna Por Reopção de Curso (13.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         376               82.0961 %
Incorrectly Classified Instances        82               17.9039 %
Kappa statistic                          0.7148

=== Confusion Matrix ===
   a   b   c   d   e   f   g   h   i   <-- classified as
 207   2   0   0   1   0   0   0   0 |   a = Aluno Regular
   3 147  10   4   0   4   7   0   0 |   b = Abandono
   0  18   7   0   0   0   1   1   0 |   c = Cancelamento
   1   5   4   4   1   0   0   1   0 |   d = Transf. Interna Por Reopção de Curso
   0   0   0   0  11   0   0   0   0 |   e = Formado
   0   5   2   1   0   0   0   0   0 |   f = Transferência
   0   5   0   0   0   0   0   0   0 |   g = Desligamento
   0   4   0   0   0   1   0   0   0 |   h = Transferência Interna
   1   0   0   0   0   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 8 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O experimento mostra dois casos distintos de desistência do curso de Engenharia de Computação. Em um cenário onde notas iguais ou inferiores a 4,6 em Calculo I e a 2,8 em Laboratório de Programação I ocorreram, 11 estudantes abandonaram seus estudos. Já em todos os casos onde houveram notas até 3,5 em

Introdução a Engenharia de Computação aliadas a notas superiores a 6,33 em Física I, os estudantes realizaram a mudança de curso dentro da própria universidade.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 9 – EC – Algoritmo J48 – Experimento 1

```
ANOS_CURSADOS <= 0
|   GEOMETRIA ANALITICA <= 2.4: Abandono (60.61/7.27)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         375               81.8777 %
Incorrectly Classified Instances        83               18.1223 %
Kappa statistic                          0.7019

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 205   2   0   1   2   0   0   0   0 |   a = Aluno Regular
   5 156   6   5   0   3   0   0   0 |   b = Abandono
   0  23   3   1   0   0   0   0   0 |   c = Cancelamento
   1  10   0   4   0   1   0   0   0 |   d = Transf. Interna Por Reopção de Curso
   4   0   0   0   7   0   0   0   0 |   e = Formado
   1   5   2   0   0   0   0   0   0 |   f = Transferência
   0   5   0   0   0   0   0   0   0 |   g = Desligamento
   0   5   0   0   0   0   0   0   0 |   h = Transferência Interna
   1   0   0   0   0   0   0   0   0 |   i = Falecimento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 9 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Computação, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Como ultimo experimento realizado no curso de Engenharia de Computação, está descrita mais uma relação entre abandono e reprovação em uma disciplina inicial. A grande maioria dos estudantes que reprovaram em Geometria Analitica, com notas iguais ou inferiores a 2,4, abandonaram a faculdade.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

4.2 Engenharia de Alimentos

Os resultados obtidos durante a mineração de dados relativa aos estudantes de Engenharia de Alimentos estão descritos nos experimentos abaixo.

Quadro 10 – EA – Algoritmo AttributeSelectedClassifier – Experimento 1

```
ANOS_CURSADOS <= 1
|   ANOS_CURSADOS <= 0: Desligamento (6.0/1.0)
|   ANOS_CURSADOS > 0: Abandono (18.0/5.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          55               63.9535 %
Incorrectly Classified Instances        31               36.0465 %
Kappa statistic                          0.5132

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
  0 10  1  2  0  0  0  0  0 |  a = Transf. Interna Por Reopção de Curso
  0 28  0  2  0  0  0  0  1 |  b = Abandono
  0  0 12  1  0  0  0  0  0 |  c = Aluno Regular
  0  0  0 10  0  0  0  0  0 |  d = Formado
  0  4  0  0  0  0  0  0  0 |  e = Transferido
  0  5  0  0  0  0  0  0  0 |  f = Transferência
  0  2  0  2  0  0  0  0  0 |  g = Cancelamento
  0  1  0  0  0  0  0  0  0 |  h = Transferência Interna
  0  0  0  0  0  0  0  0  5 |  i = Desligamento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 10 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Alimentos, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O experimento apresentado no Quadro 10 mostra a relação da quantidade de anos cursados em Engenharia de Alimentos e o abandono do curso. Os alunos que cursaram até um ano estão divididos em dois grupos: os que não completaram um ano cursado e realizaram o desligamento em 5 casos, e os alunos que completaram um ano de curso e abandonaram em 13 casos. Com base nesses dados, é possível afirmar que a evasão ocorre de maneira elevada durante o primeiro ano de curso.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como

abandono e desligamento foram avaliados de maneira correta na grande maioria dos casos.

Quadro 11 – EA – Algoritmo AttributeSelectedClassifier – Experimento 2

```
CIENCIA DOS MATERIAIS <= 2
|   FÍSICA II <= 0.7: Abandono (35.0/9.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          52               60.4651 %
Incorrectly Classified Instances        34               39.5349 %
Kappa statistic                          0.4887

=== Confusion Matrix ===

  a  b   c  d  e  f  g  h  i   <-- classified as
  1  7   1  1  0  2  1  0  0 |  a = Transf. Interna Por Reopção de Curso
  5 24  0  0  0  1  0  0  1 |  b = Abandono
  0  0 12  1  0  0  0  0  0 |  c = Aluno Regular
  0  0  0 10  0  0  0  0  0 |  d = Formado
  1  3  0   0  0  0  0  0  0 |  e = Transferido
  3  2  0   0  0  0  0  0  0 |  f = Transferência
  2  1  0   0  0  1  0  0  0 |  g = Cancelamento
  0  1  0   0  0  0  0  0  0 |  h = Transferência Interna
  0  0  0   0  0  0  0  0  5 |  i = Desligamento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 11 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Alimentos, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O estudo demonstra uma relação entre duas disciplinas básicas com o abandono. Existem 35 alunos que obtiveram notas iguais ou inferiores a 2 em Ciência dos Materiais e 0,7 em Física II combinadas. Dentre os 35, 26 evadiram, provavelmente em consequência das reprovações.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 12 – EA – Algoritmo FilteredClassifier – Experimento 1

```
ANOS_CURSADOS = '(1.5-4.5]'
|   CALCULO I = '(-inf-3.95]': Abandono (11.35/2.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          50               58.1395 %
Incorrectly Classified Instances        36               41.8605 %
Kappa statistic                          0.4662

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i   <-- classified as
  5  5  1  1  0  1  0  0  0 |  a = Transf. Interna Por Reopção de Curso
  7 20  0  2  1  0  0  0  1 |  b = Abandono
  0  0 12  1  0  0  0  0  0 |  c = Aluno Regular
  1  1  0  8  0  0  0  0  0 |  d = Formado
  1  3  0  0  0  0  0  0  0 |  e = Transferido
  3  2  0  0  0  0  0  0  0 |  f = Transferência
  0  1  0  2  1  0  0  0  0 |  g = Cancelamento
  0  1  0  0  0  0  0  0  0 |  h = Transferência Interna
  0  0  0  0  0  0  0  0  5 |  i = Desligamento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 12 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Alimentos, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O experimento mostra que, dentre os alunos que cursaram entre 1,5 e 4,5 anos, dos 11 estudantes que tiraram notas inferiores a 4 em Cálculo I, apenas 2 não abandonaram o curso. Com isso, o Quadro 12 mostra outro exemplo em que reprovações em disciplinas iniciais levam à desistência.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 13 – EA – Algoritmo FilteredClassifier – Experimento 2

```
ANOS_CURSADOS = '(1.5-4.5]'
|   FISICO-QUÍMICA II = '(-inf-2.5]'
|   |   LABORATÓRIO DE FÍSICA II = '(-inf-2.35]': Abandono (14.0/2.0)
|   |   LABORATÓRIO DE FÍSICA II = '(2.35-inf)'
|   |   |   QUÍMICA DE ALIMENTOS = '(-inf-5.55]': Transf. Interna Por Reopção de Curso (9.0/2.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          51               59.3023 %
Incorrectly Classified Instances        35               40.6977 %
Kappa statistic                          0.4739

=== Confusion Matrix ===

  a  b   c  d  e  f  g  h  i   <-- classified as
  3  7   1  0  0  1  1  0  0 |  a = Transf. Interna Por Reopção de Curso
  5 23  0  0  1  0  1  0  1 |  b = Abandono
  0  1 10  2  0  0  0  0  0 |  c = Aluno Regular
  0  0  0 10  0  0  0  0  0 |  d = Formado
  1  3  0   0  0  0  0  0  0 |  e = Transferido
  3  1  0   0  0  1  0  0  0 |  f = Transferência
  2  2  0   0  0  0  0  0  0 |  g = Cancelamento
  0  0  0   0  0  1  0  0  0 |  h = Transferência Interna
  0  1  0  0   0  0  0  0  4 |  i = Desligamento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 13 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Alimentos, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O estudo apresentado acima demonstra mais casos de abandono em consequência de reprovações. Dentre os alunos que cursaram entre 1,5 e 4,5 anos, obtiveram notas em Físico-Química II entre 0 e 2,5 e reprovação em Laboratório de Física II ocorreram 12 abandonos. Já entre os estudantes que tiraram acima de 2,35 em Física II e reprovaram em Química de Alimentos, ocorreram casos de transferências para outro curso dentro da UNIPAMPA, onde dos 9 casos de alunos que reprovaram em uma cadeira especifica do curso de Engenharia de Alimentos, apenas 2 continuaram no curso.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos e os alunos

classificados em Transferência Interna por Reopção de Curso tiveram 7 classificações corretas e 6 incorretas.

Quadro 14 – EA – Algoritmo Ridor – Experimento 1

```
(QUÍMICA GERAL TEÓRICA <= 3.65) => FORMA_EVASÃO = Abandono  (12.0/0.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          46               53.4884 %
Incorrectly Classified Instances        40               46.5116 %
Kappa statistic                          0.4173

=== Confusion Matrix ===

  a  b   c  d  e  f  g  h  i   <-- classified as
  3  5   1  2  0  2  0  0  0 |  a = Transf. Interna Por Reopção de Curso
  6 17   0  2  4  1  0  0  1 |  b = Abandono
  0  0 12  1  0  0  0  0  0 |  c = Aluno Regular
  0  2   0  8  0  0  0  0  0 |  d = Formado
  0  2   0  0  1  1  0  0  0 |  e = Transferido
  2  3   0  0  0  0  0  0  0 |  f = Transferência
  0  2   0  1  1  0  0  0  0 |  g = Cancelamento
  0  0   0  0  0  1  0  0  0 |  h = Transferência Interna
  0  0   0  0  0  0  0  0  5 |  i = Desligamento
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 14 foram gerados a partir da mineração de dados de um arquivo de Engenharia de Alimentos, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Esse experimento também expressa um relação entre abandono com reprovações em uma disciplina base do curso de Engenharia de Alimentos. Todos os estudantes que tiraram notas iguais ou inferiores a 3,65 em Química Geral Teórica acabaram desistindo de seus cursos.

A estatística de Kappa demonstra que a concordância dos resultados é razoável e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

4.3 Licenciatura em Física

Os resultados obtidos durante a mineração de dados relativa aos estudantes de Licenciatura em Física estão descritos nos experimentos abaixo.

Quadro 15 – LF – Algoritmo FilteredClassifier – Experimento 1

```
ANOS_CURSADOS = '(0.5-inf)'
|   LABORATÓRIO DE FÍSICA I = '(-inf-4.1]': Aluno Regular (37.86/12.47)
|   LABORATÓRIO DE FÍSICA I = '(4.1-inf)'
|   |   CALCULO I = '(-inf-2.35]': Abandono (14.2/6.68)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         203               63.0435 %
Incorrectly Classified Instances       119               36.9565 %
Kappa statistic                          0.4313

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 116   1   6   6   0   4   0   0 |   a = Aluno Regular
   1   0   1   0   0   3   0   0 |   b = Formado
  16   0   4  16   0   0   0   0 |   c = Cancelamento
  19   2   8  81   0   5   0   0 |   d = Abandono
   1   0   0   5   0   2   0   0 |   e = Transferência Interna
   7   0   1   7   0   2   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   1   0   1   0   0 |   g = Transferido
   4   0   0   1   0   1   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 15 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O estudo mostra uma relação entre evasão e disciplinas iniciais, porém de forma contrária aos experimentos analisados anteriormente. Entre os alunos que cursaram acima de 0,5 anos e obtiveram notas até 4,1 em Laboratório de Física I, 25 seguem como alunos regulares. Já entre os alunos que tiraram acima de 4,1 de média, ocorreram abandonos, quando aliados a reprovações em Calculo I.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados em Aluno Regular e Abandono foram classificados de maneira correta na maioria dos casos.

Quadro 16 – LF – Algoritmo JRip – Experimento 1

```
(CALCULO I <= 5.1) and (ANOS_CURSADOS <= 1) and (LABORATÓRIO DE FÍSICA I <= 2) =>
FORMA_EVASÃO=Abandono (73.0/7.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         240               74.5342 %
Incorrectly Classified Instances        82               25.4658 %
Kappa statistic                          0.6134

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 124   5   0   0   0   3   0   1 |   a = Aluno Regular
   0   5   0   0   0   0   0   0 |   b = Formado
   5   0  13  16   0   2   0   0 |   c = Cancelamento
   8   0   9  97   0   1   0   0 |   d = Abandono
   2   0   1   4   1   0   0   0 |   e = Transferência Interna
   4   0   1  11   1   0   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   1   0   1   0   0 |   g = Transferido
   1   0   3   2   0   0   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 16 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Esse experimento demonstra que a maioria alunos que cursaram até um ano de faculdade e reprovaram em Cálculo I e Laboratório de Física I, disciplinas que são iniciais do curso de Licenciatura em Física, abandonaram. Dentre os 73 alunos que se encaixam nesse perfil, apenas 7 não abandonaram os estudos.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 17 – LF – Algoritmo JRip – Experimento 2

```
(LABORATÓRIO DE FÍSICA I >= 2.6) and (ANOS_CURSADOS <= 1) and (ALGORITMOS E
PROGRAMAÇÃO <= 1.4) => FORMA_EVASAO=Cancelamento (11.0/1.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         240               74.5342 %
Incorrectly Classified Instances        82               25.4658 %
Kappa statistic                          0.6149

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 128   3   0   1   0   1   0   0 |   a = Aluno Regular
   1   4   0   0   0   0   0   0 |   b = Formado
   4   0  15  15   0   1   0   1 |   c = Cancelamento
  13   0   9  88   1   4   0   0 |   d = Abandono
   2   0   1   4   0   1   0   0 |   e = Transferência Interna
   4   0   2   7   0   4   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   1   0   1   0   0 |   g = Transferido
   0   0   3   2   0   0   0   1 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 17 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O experimento acima mostra um resultado similar ao estudo anterior, onde a evasão ocorre durante o primeiro ano de faculdade aliado a reprovações. Alunos que obtiveram notas superiores a 2,6 em Laboratório de Física I e inferiores a 1,4 em Algoritmos e Programação cancelaram o curso ainda no primeiro ano. Apenas um aluno incluso nessa classificação não realizou o cancelamento.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta 15 dos 36 casos.

Quadro 18 – LF – Algoritmo JRip – Experimento 3

```
(FUNDAMENTOS DA EDUCAÇÃO I >= 7.2) and (FÍSICA I-QT >= 2) =>
FORMA_EVASÃO=Abandono (4.0/0.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         244               75.7764 %
Incorrectly Classified Instances        78               24.2236 %
Kappa statistic                          0.6351

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 125   3   0   3   0   2   0   0 |   a = Aluno Regular
   0   4   0   1   0   0   0   0 |   b = Formado
   1   0  15  18   1   0   0   1 |   c = Cancelamento
   6   0   9  95   2   3   0   0 |   d = Abandono
   0   0   1   4   2   1   0   0 |   e = Transferência Interna
   7   0   1   6   1   2   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   2   0   0   0   0 |   g = Transferido
   1   0   3   1   0   0   0   1 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 18 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Notas” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Esse experimento demonstra mais uma ligação entre reprovações e abandono. Alunos que mesmo aprovados em Fundamentos da Educação I, uma cadeira básica em cursos voltados para a área de educação, mas que cursaram pelo menos 2 vezes Física I, abandonaram.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 19 – LF – Algoritmo JRip – Experimento 4

```
(LABORATÓRIO DE FÍSICA I = Aprovado com nota) and (HISTÓRIA DA EDUCAÇÃO = 0.0) and
(GEOMETRIA ANALITICA = Aprovado com nota) => FORMA_EVASÃO=Transf. Interna Por
Reopção de Curso (8.0/2.0)

(FUNDAMENTOS DA EDUCAÇÃO I = Reprovado com nota) => FORMA_EVASÃO=Transf.
Interna Por Reopção de Curso (3.0/1.0)

(QUÍMICA ORGÂNICA = Reprovado por Frequência) => FORMA_EVASÃO=Abandono (4.0/0.0)

(LABORATÓRIO DE FÍSICA I = Reprovado por Frequência) and (ANOS_CURSADOS <= 1) =>
FORMA_EVASÃO=Abandono (15.0/2.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         239               74.2236 %
Incorrectly Classified Instances        83               25.7764 %
Kappa statistic                          0.607

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 128   3   0   0   0   2   0   0 |   a = Aluno Regular
   2   3   0   0   0   0   0   0 |   b = Formado
   7   0  18  10   0   1   0   0 |   c = Cancelamento
  11   1   9  90   1   3   0   0 |   d = Abandono
   3   0   1   3   0   1   0   0 |   e = Transferência Interna
   9   0   2   6   0   0   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   2   0   0   0   0 |   g = Transferido
   1   0   4   1   0   0   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 19 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Aproveitamento” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

As duas primeiras partes do experimento demonstram que alunos que obtiveram aprovação em cadeiras presentes em outros cursos dentro da UNIPAMPA, como Laboratório de Física I e Geometria Analitica, aliados a reprovações em cadeiras presentes em cursos da área da educação, como História da Educação e Fundamentos de Educação I, realizaram a reopção de curso. As duas ultimas partes do experimento mostram mais casos de abandono do curso, onde reprovações por frequência em cadeiras básicas – Química Orgânica e

Laboratório de Física I – levaram ao abandono. No caso de Laboratório de Física, as desistências ocorreram durante o primeiro ano de curso.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos, porém os alunos classificados como transferência interna por reopção de curso apresentaram classificação errônea.

Quadro 20 – LF – Algoritmo JRip – Experimento 5

```
(FÍSICA II-QT >= 2) and (ANO_INGRESSO <= 2006) => FORMA_EVASÃO=Transferência Interna
(5.0/2.0)

(LABORATÓRIO DE FÍSICA I = Aprovado com nota) and (ANO_INGRESSO <= 2008) and
(CALCULO I-QT >= 2) => FORMA_EVASÃO=Transf. Interna Por Reopção de Curso (12.0/4.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         245               76.087  %
Incorrectly Classified Instances        77               23.913  %
Kappa statistic                          0.6386

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 126   3   0   2   0   2   0   0 |   a = Aluno Regular
   0   3   0   1   0   1   0   0 |   b = Formado
   8   0  15  11   1   1   0   0 |   c = Cancelamento
   4   0  10  98   1   2   0   0 |   d = Abandono
   2   0   1   2   0   3   0   0 |   e = Transferência Interna
   5   0   2   7   0   3   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   1   0   1   0   0 |   g = Transferido
   2   0   3   1   0   0   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 20 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Aproveitamento” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O estudo acima mostra que 3 alunos que cursaram pelo menos duas vezes Física II e ingressaram na faculdade até 2006 realizaram a reopção de curso. O mesmo ocorre com alunos que cursaram Cálculo I pelo menos duas vezes e foram

aprovados em Laboratório de Física I, entrando no curso até 2008, onde ocorreram 8 casos de reopção.

A estatística de Kappa demonstra que a concordância dos resultados é substancial, porém a matriz de confusão apresentou uma classificação errônea na maioria dos casos classificados como transferência interna por reopção de curso.

Quadro 21 – LF – Algoritmo PART – Experimento 1

```
ANO_INGRESSO <= 2012 AND
ANOS_CURSADOS <= 1 AND
LABORATÓRIO DE FÍSICA I <= 4.3: Abandono (92.67/13.49)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         262               81.3665 %
Incorrectly Classified Instances        60               18.6335 %
Kappa statistic                          0.7183

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 132   0   0   0   0   1   0   0 |   a = Aluno Regular
   1   3   0   0   1   0   0   0 |   b = Formado
   0   0  13  18   0   3   0   2 |   c = Cancelamento
   1   0   7 105   0   2   0   0 |   d = Abandono
   0   2   0   5   1   0   0   0 |   e = Transferência Interna
   0   0   1   8   0   8   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   1   0   1   0   0 |   g = Transferido
   0   0   3   3   0   0   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 21 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

A mineração de dados demonstrada acima mostra mais um relação com reprovações em disciplinas iniciais e abandono. Entre os 92 alunos que ingressaram no curso de Licenciatura em Física até o ano de 2012, cursaram até um ano de faculdade e obtiveram notas inferiores a 4,4 em Laboratório de Física I, apenas 13 não abandonaram a faculdade.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 22 – LF – Algoritmo PART – Experimento 2

```
ANO_INGRESSO <= 2008 AND
ANOS_CURSADOS <= 3 AND
ORGANIZAÇÃO ESCOLAR E TRABALHO DOCENTE = Reprovado por Frequência: Abandono
(14.3/3.89)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         258               80.1242 %
Incorrectly Classified Instances        64               19.8758 %
Kappa statistic                          0.6979

=== Confusion Matrix ===

  a   b   c   d   e   f   g   h   <-- classified as
132   0   0   0   0   1   0   0 |   a = Aluno Regular
  1   3   0   1   0   0   0   0 |   b = Formado
  0   0  17  17   0   2   0   0 |   c = Cancelamento
  1   1   7 102   1   3   0   0 |   d = Abandono
  0   2   0   6   0   0   0   0 |   e = Transferência Interna
  0   0   0  13   0   4   0   0 |   f = Transf. Interna Por Reopção de Curso
  0   0   0   2   0   0   0   0 |   g = Transferido
  0   0   3   2   0   1   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 22 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Aproveitamento” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O estudo acima mostra que, entre os alunos que entraram antes de 2008 na UNIPAMPA, os estudantes que cursaram até três anos de faculdade e desistiram de cursar a disciplina Organização Escolar e Trabalho Docente acabaram abandonando os estudos na instituição.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 23 – LF – Algoritmo PART – Experimento 3

```
LABORATÓRIO DE FÍSICA II = Aprovado com nota AND
INSTRUMENTAÇÃO PARA O ENSINO DE FÍSICA I = 0.0: Transf. Interna Por Reopção de Curso
(9.0/2.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         235               72.9814 %
Incorrectly Classified Instances        87               27.0186 %
Kappa statistic                          0.5982

=== Confusion Matrix ===

  a   b   c   d   e   f   g   h   <-- classified as
128   2   1   1   0   0   0   1 |   a = Aluno Regular
  3   2   0   0   0   0   0   0 |   b = Formado
  1   0  10  18   1   1   0   5 |   c = Cancelamento
  1   0  14  93   0   5   0   2 |   d = Abandono
  1   0   1   3   0   3   0   0 |   e = Transferência Interna
  2   0   2   7   3   2   0   1 |   f = Transf. Interna Por Reopção de Curso
  0   0   0   1   0   1   0   0 |   g = Transferido
  0   0   1   3   0   2   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 23 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “0” e “Aproveitamento” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

No experimento acima pode-se verificar mais um provável perfil presente entre os estudantes que realizam reopção de curso dentro da UNIPAMPA. Entre os 9 alunos que foram aprovados em Laboratório de Física II e não obtiveram aprovação em Instrumentação para o Ensino de Física I, 7 realizaram a reopção para algum outro curso dentro da universidade.

A estatística de Kappa demonstra que a concordância dos resultados é moderada, porém a matriz de confusão apresentou uma classificação errônea na maioria dos casos classificados como transferência interna por reopção de curso.

Quadro 24 – LF – Algoritmo J48 – Experimento 1

```
ANO_INGRESSO <= 2012
|   ANOS_CURSADOS <= 1
|   |   LABORATÓRIO DE FÍSICA I <= 4.3: Abandono (92.67/13.49)
|   |   LABORATÓRIO DE FÍSICA I > 4.3
|   |   |   FÍSICA I <= 7.3: Cancelamento (39.03/16.62)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         253               78.5714 %
Incorrectly Classified Instances        69               21.4286 %
Kappa statistic                          0.6734

=== Confusion Matrix ===
   a   b   c   d   e   f   g   h   <-- classified as
 132   0   0   0   0   1   0   0 |   a = Aluno Regular
   0   3   0   1   1   0   0   0 |   b = Formado
   1   0  14  20   0   1   0   0 |   c = Cancelamento
   1   0   9 101   0   3   0   1 |   d = Abandono
   0   2   0   5   1   0   0   0 |   e = Transferência Interna
   4   0   1  10   0   2   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   1   1   0   0   0   0 |   g = Transferido
   0   0   4   2   0   0   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 24 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados com “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O estudo mostra que, dentre os alunos que entraram antes de 2013 e cursaram até um ano, aqueles que reprovaram em Laboratório de Física I abandonaram na grande maioria dos casos. Já entre os alunos que tiraram notas superiores a 4,3 em Laboratório de Física I e menores ou iguais a 7,3 em Física I, ocorreram 23 cancelamentos.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono e cancelamento foram avaliados de maneira correta na grande maioria dos casos.

Quadro 25 – LF – Algoritmo J48 – Experimento 2

```
ANOS_CURSADOS > 0
|   LABORATÓRIO DE FÍSICA I-QT <= 1
|   |   FÍSICA II-QT > 1: Abandono (55.0/21.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         242               75.1553 %
Incorrectly Classified Instances        80               24.8447 %
Kappa statistic                          0.6265

=== Confusion Matrix ===
  a   b   c   d   e   f   g   h   <-- classified as
 130  2   0   0   1   0   0   0 |   a = Aluno Regular
   1   3   0   1   0   0   0   0 |   b = Formado
   1   0  11  23   1   0   0   0 |   c = Cancelamento
   1   1  16  93   0   4   0   0 |   d = Abandono
   1   0   1   4   1   1   0   0 |   e = Transferência Interna
   2   1   2   7   1   4   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   1   0   1   0   0 |   g = Transferido
   1   0   3   2   0   0   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 25 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O estudo mostra que entre os alunos que cursaram pelo menos um ano de faculdade, no máximo uma vez a disciplina de Laboratório de Física I e pelo menos duas vezes a disciplina de Física II, ocorreram 34 abandonos entre os 55 estudantes que se enquadram nessa situação.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 26 – LF – Algoritmo J48 – Experimento 3

```
QUÍMICA GERAL <= 6.1
|   CALCULO I <= 0.7: Abandono (83.0/9.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         242               75.1553 %
Incorrectly Classified Instances        80               24.8447 %
Kappa statistic                          0.6296

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
 127   6   0   0   0   0   0   0 |   a = Aluno Regular
   1   2   1   0   1   0   0   0 |   b = Formado
   1   0  11  21   1   2   0   0 |   c = Cancelamento
   1   0  12  95   0   4   2   1 |   d = Abandono
   1   0   1   3   2   1   0   0 |   e = Transferência Interna
   4   0   2   6   0   5   0   0 |   f = Transf. Interna Por Reopção de Curso
   0   0   0   1   0   1   0   0 |   g = Transferido
   0   0   2   3   1   0   0   0 |   h = Transferência
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 26 foram gerados a partir da mineração de dados de um arquivo de Licenciatura em Física, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Como ultimo experimento do curso de Licenciatura em Física, este estudo mostra mais uma relação com disciplinas iniciais e abandono. Dentre os 83 alunos que obtiveram notas inferiores a 6,2 em Química Geral e 0,8 em Cálculo I, apenas 9 não abandonaram o curso.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

4.4 Engenharia de Produção

Os resultados obtidos durante a mineração de dados relativa aos estudantes de Engenharia de Produção estão descritos nos experimentos abaixo.

Quadro 27 – EP – Algoritmo FilteredClassifier – Experimento 1

```
ANOS_CURSADOS = '(-inf-0.5]'
|   PRODUÇÃO ACADEMICO CIENTÍFICA = '(-inf-5.35]': Abandono (51.05/9.38)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         420               77.7778 %
Incorrectly Classified Instances       120               22.2222 %
Kappa statistic                          0.6189

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
   6   0  21   0   0   0   0   0 |   a = Formado
   0 171  22   0   1   0   1   0 |   b = Abandono
   4  19 238   0   0   0   0   0 |   c = Aluno Regular
   0   6   5   0   0   0   0   0 |   d = Transferência
   0  11   1   0   0   0   0   0 |   e = Transf. Interna Por Reopção de Curso
   0   4   0   0   0   0   0   0 |   f = Desligamento
   0  20   3   0   0   0   5   0 |   g = Cancelamento
   0   2   0   0   0   0   0   0 |   h = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 27 foram gerados a partir da mineração de dados de um arquivo Engenharia de Produção, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Como primeiro experimento do curso de Engenharia de Produção, é demonstrada acima uma relação entre abandono e reprovação em disciplinas iniciais. Dentre os 51 alunos que cursaram menos de um ano de faculdade, somente 9 tiraram notas superiores a 5,35 em Produção Acadêmico Cientifica.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 28 – EP – Algoritmo AttributeSelectedClassifier – Experimento 1

```
ANOS_CURSADOS > 0
|   ANO_INGRESSO <= 2010
|   |   ANOS_CURSADOS <= 3
|   |   |   ENGENHARIA ECONÔMICA II <= 6: Abandono (135.0/28.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         404               74.8148 %
Incorrectly Classified Instances       136               25.1852 %
Kappa statistic                          0.5679

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
  27   0   0   0   0   0   0   0 |   a = Formado
   0 126  67   0   0   1   1   0 |   b = Abandono
   3  10 248   0   0   0   0   0 |   c = Aluno Regular
   0   6   5   0   0   0   0   0 |   d = Transferência
   0   8   4   0   0   0   0   0 |   e = Transf. Interna Por Reopção de Curso
   0   3   1   0   0   0   0   0 |   f = Desligamento
   0  14  11   0   0   0   3   0 |   g = Cancelamento
   0   1   1   0   0   0   0   0 |   h = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 28 foram gerados a partir da mineração de dados de um arquivo Engenharia de Produção, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Entre os alunos que cursaram de um a três anos e ingressaram até 2010 no curso de Engenharia de Produção, existem 135 casos de estudantes que tiraram notas iguais ou menores que 6 em Engenharia Econômica II. Ocorreram 107 casos de abandono entre os 135 que se encaixam nessa classificação, demonstrando que a cadeira de Engenharia Econômica II possui uma grande ligação com a evasão.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 29 – EP – Algoritmo JRip – Experimento 1

```
(CALCULO I <= 3) and (ANO_INGRESSO <= 2010) and (ANOS_CURSADOS <= 3) =>
FORMA_EVASÃO=Abandono (101.0/12.0)

(SISTEMAS PRODUTIVOS I <= 4.74) and (ANOS_CURSADOS <= 1) and (ANO_INGRESSO <=
2012) => FORMA_EVASÃO=Abandono (40.0/3.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         450               83.3333 %
Incorrectly Classified Instances        90               16.6667 %
Kappa statistic                          0.7225

=== Confusion Matrix ===

  a   b   c   d   e   f   g   h   <-- classified as
 26   1   0   0   0   0   0   0 |   a = Formado
  0 171  19   0   0   0   5   0 |   b = Abandono
  3   6 250   2   0   0   0   0 |   c = Aluno Regular
  0   9   1   0   0   0   1   0 |   d = Transferência
  0  10   2   0   0   0   0   0 |   e = Transf. Interna Por Reopção de Curso
  0   4   0   0   0   0   0   0 |   f = Desligamento
  0  22   2   1   0   0   3   0 |   g = Cancelamento
  0   2   0   0   0   0   0   0 |   h = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 29 foram gerados a partir da mineração de dados de um arquivo Engenharia de Produção, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

No Quadro 29 estão listados mais dois casos onde reprovações em disciplinas iniciais podem ter sido causa de evasão. Entre os 101 casos de alunos que obtiveram notas iguais ou menores a 3 em Calculo I, cursaram até três anos e ingressaram no curso até 2010, ocorreram 89 abandonos. Já entre os 40 estudantes que obtiveram notas iguais ou inferiores a 4,74 em Sistemas Produtivos I, estudaram até completar um ano de curso e ingressaram até o ano de 2012, apenas 3 não abandonaram a universidade.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 30 – EP – Algoritmo JRip – Experimento 2

```
(LABORATORIO DE FÍSICA I <= 3.5) and (ANO_INGRESSO <= 2010) and (SISTEMAS
PRODUTIVOS II <= 0) => FORMA_EVASÃO=Abandono (121.0/20.0)

(ECONOMIA INDUSTRIAL <= 5.5) and (ANO_INGRESSO <= 2012) and (ANOS_CURSADOS <=
1) => FORMA_EVASÃO=Abandono (57.0/13.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         446               82.5926 %
Incorrectly Classified Instances        94               17.4074 %
Kappa statistic                          0.7085

=== Confusion Matrix ===

  a   b   c   d   e   f   g   h   <-- classified as
 26   1   0   0   0   0   0   0 |   a = Formado
  0 169  22   0   0   0   4   0 |   b = Abandono
  3   7 249   2   0   0   0   0 |   c = Aluno Regular
  0   5   5   0   0   0   1   0 |   d = Transferência
  0  10   2   0   0   0   0   0 |   e = Transf. Interna Por Reopção de Curso
  0   4   0   0   0   0   0   0 |   f = Desligamento
  0  23   3   0   0   0   2   0 |   g = Cancelamento
  0   2   0   0   0   0   0   0 |   h = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 30 foram gerados a partir da mineração de dados de um arquivo Engenharia de Produção, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

No experimento acima podem ser observados mais dois casos onde reprovações em disciplinas iniciais podem ter sido causa de evasão. Na primeira parte do experimento pode-se observar que entre os 121 alunos que reprovaram em Laboratório de Física I, tendo ingressado na faculdade até 2010 e que não obtiveram nota em Sistemas Produtivos II, ocorreram 101 casos de abandono. Na segunda parte está expressa uma relação com Economia Industrial, onde dos 57 alunos que obtiveram notas iguais ou inferiores a 5,5 na disciplina, tendo cursado até um ano de curso e entrado na universidade antes de 2013, apenas 13 não abandonaram.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 31 – EP – Algoritmo PART – Experimento 1

```
ANOS_CURSADOS <= 1 AND
SISTEMAS PRODUTIVOS II <= 2: Abandono (85.85/17.46)

ANOS_CURSADOS <= 1 AND
FORMA_INGRESSO = Portador de Diploma: Abandono (9.75/0.38)

FORMA_INGRESSO = Processo Seletivo - Vestibular AND
DESENHO TÉCNICO II <= 9 AND
ERGONOMIA II <= 3.15: Abandono (12.71/1.56)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         466               86.2963 %
Incorrectly Classified Instances        74               13.7037 %
Kappa statistic                          0.7742

=== Confusion Matrix ===

  a   b   c   d   e   f   g   h   <-- classified as
 22   3   1   1   0   0   0   0 |   a = Formado
  5 179   3   3   4   0   1   0 |   b = Abandono
  1   0 260   0   0   0   0   0 |   c = Aluno Regular
  2   6   0   2   1   0   0   0 |   d = Transferência
  2   7   2   1   0   0   0   0 |   e = Transf. Interna Por Reopção de Curso
  0   4   0   0   0   0   0   0 |   f = Desligamento
  0  25   0   0   0   0   3   0 |   g = Cancelamento
  0   2   0   0   0   0   0   0 |   h = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 31 foram gerados a partir da mineração de dados de um arquivo Engenharia de Produção, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Mais casos de abandono podem ser observados no Quadro 31. Ocorreram 58 casos de abandono entre os alunos que cursaram até um ano de Engenharia de Produção e obtiveram notas iguais ou inferiores a 2 em Sistemas Produtivos II. Entre os estudantes que ingressaram no curso como Portadores de Diploma e cursaram até um ano de curso, todos os 9 evadiram. Já entre os alunos que ingressaram através do Vestibular e obtiveram notas iguais ou inferiores a 9 em Desenho Técnico II e 3,15 em Ergonomia II, ocorreram 11 casos de abandono.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 32 – EP – Algoritmo PART – Experimento 2

```
ANO_INGRESSO > 2007 AND
FORMA_INGRESSO = Transferência Voluntária ou Externa (oriundo de outra instituição):
Abandono (6.0/1.0)

ANOS_CURSADOS <= 5 AND
FORMA_INGRESSO = Transferência EX-OFFICIO (amparada em lei): Abandono (4.0/1.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         460               85.1852 %
Incorrectly Classified Instances        80               14.8148 %
Kappa statistic                          0.7547

=== Confusion Matrix ===

  a   b   c   d   e   f   g   h   <-- classified as
 22   0   5   0   0   0   0   0 |   a = Formado
  9 175   4   0   2   0   5   0 |   b = Abandono
  1   1 259   0   0   0   0   0 |   c = Aluno Regular
  1   8   1   0   0   0   1   0 |   d = Transferência
  1   8   2   0   1   0   0   0 |   e = Transf. Interna Por Reopção de Curso
  0   4   0   0   0   0   0   0 |   f = Desligamento
  0  24   1   0   0   0   3   0 |   g = Cancelamento
  0   2   0   0   0   0   0   0 |   h = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 32 foram gerados a partir da mineração de dados de um arquivo Engenharia de Produção, onde foi utilizada a configuração de pré-processamento de dados “?” e “Aproveitamento” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

No experimento acima são listados mais casos de abandono relacionados a forma de ingresso na universidade. Entre os 6 alunos que utilizaram Transferência Voluntária ou Externa como meio de entrada no curso ocorreram 5 abandonos. Já entre os 4 estudantes que utilizaram Transferência Ex-Officio, ocorreram 3 casos de evasão.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 33 – EP – Algoritmo J48 – Experimento 1

```
ANOS_CURSADOS <= 3
|   GESTÃO DA QUALIDADE II <= 6.3
|   |   CALCULO II <= 6.1: Abandono (121.0/16.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         469               86.8519 %
Incorrectly Classified Instances        71               13.1481 %
Kappa statistic                          0.7851

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   <-- classified as
  27   0   0   0   0   0   0   0 |   a = Formado
   0 177   9   1   1   0   7   0 |   b = Abandono
   3   2 255   0   1   0   0   0 |   c = Aluno Regular
   0   5   1   1   2   0   2   0 |   d = Transferência
   0   8   1   1   2   0   0   0 |   e = Transf. Interna Por Reopção de Curso
   0   4   0   0   0   0   0   0 |   f = Desligamento
   0  18   1   1   1   0   7   0 |   g = Cancelamento
   0   2   0   0   0   0   0   0 |   h = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 33 foram gerados a partir da mineração de dados de um arquivo Engenharia de Produção, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Como ultimo experimento de Engenharia de Produção, é listado mais um caso de reprovações em disciplinas iniciais. Ocorreram 105 abandonos entre os 121 alunos que cursaram até três anos e obtiveram notas iguais ou inferiores a 6,4 em Gestão da Qualidade II e 6,1 em Calculo II.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

4.5 Engenharia Química

Os resultados obtidos durante a mineração de dados relativa aos estudantes de Engenharia Química estão descritos nos experimentos abaixo.

Quadro 34 – EQ – Algoritmo AttributeSelectedClassifier – Experimento 1

```
ANO_INGRESSO <= 2010
|   ANOS_CURSADOS <= 3
|   |   PROBABILIDADE E ESTATISTICA <= 0.1
|   |   |   GEOMETRIA ANALITICA <= 4.8: Abandono (46.0/12.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         331               78.8095 %
Incorrectly Classified Instances        89               21.1905 %
Kappa statistic                          0.6596

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 219   0   4   0   5   0   1   0   0 |   a = Aluno Regular
   1   1   0   0   3   0   1   0   0 |   b = Transferido
   1   0  36   0   0   0   0   0   0 |   c = Formado
   3   0   0   0   7   0   2   1   0 |   d = Transf. Interna Por Reopção de Curso
  12   2   0   1  68   0   7   5   0 |   e = Abandono
   0   0   0   0   2   0   0   0   0 |   f = Desligamento
   2   1   0   3  10   0   3   2   0 |   g = Cancelamento
   1   1   0   2   6   0   1   4   0 |   h = Transferência
   0   0   0   0   2   0   0   0   0 |   i = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 34 foram gerados a partir da mineração de dados de um arquivo Engenharia Química, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Ocorreram 34 casos de abandonos entre os 46 alunos de Engenharia Química que ingressaram antes de 2011, cursaram até três anos e obtiveram notas menores ou iguais a 0,1 em Probabilidade e Estatística e 4,8 em Geometria Analitica. Isso demonstra uma relação entre reprovações em disciplinas base dos cursos de Engenharia e abandonos.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

Quadro 35 – EQ – Algoritmo FilteredClassifier – Experimento 1

```
ANOS_CURSADOS = '(0.5-3.5]'
|   FORMA_INGRESSO = Processo Seletivo - Vestibular
|   |   QUÍMICA GERAL EXPERIMENTAL = '(-inf-4.55]': Abandono (19.0/3.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         317               75.4762 %
Incorrectly Classified Instances       103               24.5238 %
Kappa statistic                          0.5958

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 217   0   2   0   8   0   1   1   0 |   a = Aluno Regular
   0   2   0   1   3   0   0   0   0 |   b = Transferido
   2   0  35   0   0   0   0   0   0 |   c = Formado
   5   1   0   0   3   0   2   2   0 |   d = Transf. Interna Por Reopção de Curso
  20   5   3   4  62   0   1   0   0 |   e = Abandono
   0   0   0   0   2   0   0   0   0 |   f = Desligamento
   4   2   1   1  11   0   1   1   0 |   g = Cancelamento
   8   0   1   4   2   0   0   0   0 |   h = Transferência
   0   0   0   0   2   0   0   0   0 |   i = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 35 foram gerados a partir da mineração de dados de um arquivo Engenharia Química, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Os alunos de Engenharia Química que cursaram entre 0,5 e 3,5 anos de faculdade, ingressaram no curso através do vestibular e tiveram como nota final em Química Geral Experimental valores entre 0 e 4,55 apresentaram um perfil propenso à evasão. Ocorreram 16 casos de abandono entre os 19 alunos que se enquadram no perfil descrito acima.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 36 – EQ – Algoritmo JRip – Experimento 1

```
(ALGEBRA LINEAR E GEOMETRIA ANALITICA <= 4.3) and (FÍSICA II <= 5.3) =>
FORMA_EVASÃO=Cancelamento (6.0/1.0)

(QUÍMICA GERAL EXPERIMENTAL <= 4.5) and (ANO_INGRESSO <= 2011) =>
FORMA_EVASÃO=Abandono (52.0/10.0)

(FÍSICA I <= 3.3) and (ANO_INGRESSO <= 2012) and (ANOS_CURSADOS <= 1) =>
FORMA_EVASÃO=Abandono (21.0/2.0)

(INTRODUÇÃO A ENGENHARIA QUÍMICA <= 3.2) and (ANO_INGRESSO <= 2011) and
(ANOS_CURSADOS <= 2) => FORMA_EVASÃO=Abandono (7.0/0.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         306               72.8571 %
Incorrectly Classified Instances       114               27.1429 %
Kappa statistic                          0.5364

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 213   0   6   0   5   0   2   3   0 |   a = Aluno Regular
   2   3   0   0   0   0   1   0   0 |   b = Transferido
   1   0  36   0   0   0   0   0   0 |   c = Formado
   8   0   0   0   3   0   2   0   0 |   d = Transf. Interna Por Reopção de Curso
  35   2   0   1  54   0   3   0   0 |   e = Abandono
   0   0   0   0   2   0   0   0   0 |   f = Desligamento
   9   0   0   1  10   0   0   1   0 |   g = Cancelamento
  10   0   0   0   4   0   1   0   0 |   h = Transferência
   0   0   0   0   2   0   0   0   0 |   i = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 36 foram gerados a partir da mineração de dados de um arquivo Engenharia Química, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

O Quadro 36 demonstra mais casos de abandono associados a reprovações em disciplinas introdutórias. Ocorreram 5 casos de cancelamento entre 6 alunos que reprovaram em Álgebra Linear e Geometria Analitica e Física II, com notas inferiores a 4,4 e 5,4, respectivamente. Entre os 52 alunos que ingressaram até 2011 em Engenharia Química e reprovaram em Química Geral Experimental com notas variando entre 0 e 4,5, ocorreram 42 casos de evasão. Já entre os 21 alunos que ingressaram no curso até 2012 e que cursaram até um ano com notas inferiores a 4,4 em Física I, apenas 2 seguiram seus estudos. Como ultimo caso listado, todos os 7 alunos que reprovaram em Introdução a Engenharia Química com notas iguais

ou inferiores a 3,2, onde ingressaram até o ano de 2011 e cursaram até dois anos abandonaram.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos e os classificados como cancelamento obtiveram alguns erros de classificação.

Quadro 37 – EQ – Algoritmo PART – Experimento 1

```
ANOS_CURSADOS <= 0 AND
ANO_INGRESSO > 2007 AND
LABORATORIO DE FÍSICA I <= 7.5: Abandono (33.43/1.93)

INTRODUÇÃO A ENGENHARIA QUÍMICA <= 5: Abandono (28.91/6.09)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         338               80.4762 %
Incorrectly Classified Instances        82               19.5238 %
Kappa statistic                          0.6862

=== Confusion Matrix ===

   a   b   c   d   e   f   g   h   i   <-- classified as
 226   0   1   0   1   0   1   0   0 |   a = Aluno Regular
   0   2   0   0   4   0   0   0   0 |   b = Transferido
   2   0  33   0   1   0   1   0   0 |   c = Formado
   0   1   1   1   8   0   1   1   0 |   d = Transf. Interna Por Reopção de Curso
   6   5   4   3  72   0   4   1   0 |   e = Abandono
   0   0   0   0   2   0   0   0   0 |   f = Desligamento
   2   0   4   0  13   0   1   1   0 |   g = Cancelamento
   0   0   2   2   8   0   0   3   0 |   h = Transferência
   0   0   0   0   2   0   0   0   0 |   i = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 37 foram gerados a partir da mineração de dados de um arquivo Engenharia Química, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Entre os alunos que ingressaram na universidade a partir de 2008 e cursaram menos de um ano, existem 33 alunos que obtiveram notas iguais ou inferiores a 7,5 em Laboratório de Física I, onde 31 evadiram. Já entre os 28 casos de estudantes que reprovaram em Introdução a Engenharia Química com notas até 5, apenas 6 não abandonaram o curso.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 38 – EQ – Algoritmo PART – Experimento 2

```
POLITICAS PUBLICAS EDUCACIONAIS NO CONTEXTO BRASILEIRO <= 7.3 AND
SEXO = F AND
HIGIENE E SEGURANÇA DO TRABALHO <= 6.8 AND
ALGORITMOS E PROGRAMAÇÃO <= 0.5: Abandono (21.0/3.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         326               77.619  %
Incorrectly Classified Instances        94               22.381  %
Kappa statistic                          0.6483

=== Confusion Matrix ===

  a   b   c   d   e   f   g   h   i   <-- classified as
212   0   6   0   5   0   4   2   0 |   a = Aluno Regular
  0   3   0   0   2   0   1   0   0 |   b = Transferido
  1   0  35   0   1   0   0   0   0 |   c = Formado
  1   2   0   1   6   0   1   2   0 |   d = Transf. Interna Por Reopção de Curso
  8   5   1   2  69   1   8   1   0 |   e = Abandono
  0   0   0   0   2   0   0   0   0 |   f = Desligamento
  3   1   0   4  10   0   1   2   0 |   g = Cancelamento
  4   0   0   0   2   0   4   5   0 |   h = Transferência
  0   0   0   0   2   0   0   0   0 |   i = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 38 foram gerados a partir da mineração de dados de um arquivo Engenharia Química, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Como ultimo experimento de Engenharia Química, é demonstrado um perfil existente entre 18 alunos evadidos. Todas as alunas do sexo feminino que obtiveram notas iguais ou inferiores a 7,3 em Politicas Publicas Educacionais no Contexto Brasileiro, 6,8 em Higiene e Segurança do Trabalho e 0,5 em Algoritmos e Programação evadiram.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na maioria dos casos.

4.6 Engenharia de Energias Renováveis e de Ambiente

Os resultados obtidos durante a mineração de dados relativa aos estudantes de Engenharia de Energias Renováveis e de Ambiente estão descritos nos experimentos abaixo.

Quadro 39 – EE – Algoritmo FilteredClassifier – Experimento 1

```
ANOS_CURSADOS = '(-inf-0.5]'
|   GEOMETRIA ANALITICA = '(-inf-2.55]': Abandono (36.21/3.68)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         292               73      %
Incorrectly Classified Instances       108               27      %
Kappa statistic                          0.5763

=== Confusion Matrix ===

   a   b   c   d   e   f   g   <-- classified as
  15   6   0   0   0   0   0 |   a = Formado
   7 160   1   1   9   0   0 |   b = Aluno Regular
   1  15   0   0   6   0   0 |   c = Transf. Interna Por Reopção de Curso
   1   7   0   0   6   0   0 |   d = Transferência
   1  16   1   0 109   8   0 |   e = Abandono
   0   2   0   0  15   8   0 |   f = Cancelamento
   0   3   0   0   2   0   0 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 39 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Como percebido em outros cursos, reprovações em Geometria Analitica também aparecem durante os experimentos de Engenharia de Energias Renováveis e de Ambiente entre os perfis de estudantes que abandonaram a universidade. Entre os estudantes que cursaram menos de um ano e não obtiveram notas superiores a 2,55 em Geometria Analitica ocorreram 33 abandonos.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 40 – EE – Algoritmo FilteredClassifier – Experimento 2

```
ANOS_CURSADOS = '(-inf-0.5]'
|   QUÍMICA GERAL = Reprovado por Frequência: Abandono (24.57/1.14)
|   QUÍMICA GERAL = Reprovado com nota: Abandono (10.75/3.5)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         280               70      %
Incorrectly Classified Instances       120               30      %
Kappa statistic                          0.5262

=== Confusion Matrix ===

   a   b   c   d   e   f   g   <-- classified as
  16   5   0   0   0   0   0 |   a = Formado
   6 153   5   0  13   0   1 |   b = Aluno Regular
   1  16   1   0   4   0   0 |   c = Transf. Interna Por Reopção de Curso
   0   9   0   0   5   0   0 |   d = Transferência
   0  20   1   0 109   2   3 |   e = Abandono
   0   2   0   0  22   1   0 |   f = Cancelamento
   0   3   0   0   2   0   0 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 40 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “?” e “Aproveitamento” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Reprovações em Química Geral também aparecem como possíveis causas da evasão. Entre os estudantes que cursaram menos de um ano e reprovaram por frequência na disciplinas, ocorreram 23 abandonos. Já entre os alunos que reprovaram por nota, ocorreram 7 abandonos.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 41 – EE – Algoritmo JRip – Experimento 1

```
(FÍSICA I <= 1.75) and (ANO_INGRESSO <= 2011) => FORMA_EVASÃO=Abandono
(100.0/17.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         285               71.25  %
Incorrectly Classified Instances       115               28.75  %
Kappa statistic                          0.5521

=== Confusion Matrix ===

  a   b   c   d   e   f   g   <-- classified as
 18   1   1   0   0   0   1 |   a = Formado
  6 154   2   3  11   0   2 |   b = Aluno Regular
  0  13   3   0   5   0   1 |   c = Transf. Interna Por Reopção de Curso
  0   5   4   0   5   0   0 |   d = Transferência
  0  24   1   0 107   3   0 |   e = Abandono
  0   6   0   0  16   3   0 |   f = Cancelamento
  0   2   0   0   3   0   0 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 41 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

A quantidade de alunos que reprovaram em Física I com ano de ingresso anterior a 2012, onde como consequência ocorreu o abandono do curso é bem alta. Entre os 100 alunos presentes na classificação, apenas 17 deram continuidade aos estudos no curso de Engenharia de Energias Renováveis e de Ambiente.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 42 – EE – Algoritmo JRip – Experimento 2

```
(FUNDAMENTOS DE ADMINISTRAÇÃO = Reprovado por Frequência) =>
FORMA_EVASÃO=Abandono (13.0/5.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         302               75.5    %
Incorrectly Classified Instances        98               24.5    %
Kappa statistic                          0.6233

=== Confusion Matrix ===

  a   b   c   d   e   f   g   <-- classified as
 20   0   0   0   0   0   1 |   a = Formado
  1 158   5   0  12   0   2 |   b = Aluno Regular
  0  10   9   0   3   0   0 |   c = Transf. Interna Por Reopção de Curso
  0   8   3   0   1   0   2 |   d = Transferência
  0  16   4   0 112   1   2 |   e = Abandono
  0   4   2   0  17   2   0 |   f = Cancelamento
  0   0   1   0   3   0   1 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 42 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “?” e “Aproveitamento” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Reprovações por frequência em Fundamentos de Administração também aparecem ligadas à evasão. Em 13 casos de estudantes que reprovaram na disciplina por frequência ocorreram 8 casos de abandono do curso.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 43 – EE – Algoritmo JRip – Experimento 3

```
(QUÍMICA GERAL EXPERIMENTAL <= 3) and (ANO_INGRESSO <= 2009) =>
FORMA_EVASÃO=Abandono (65.0/7.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         291               72.75  %
Incorrectly Classified Instances       109               27.25  %
Kappa statistic                          0.5773

=== Confusion Matrix ===

  a   b   c   d   e   f   g   <-- classified as
 18   1   0   0   0   0   2 |   a = Formado
  7 156   6   0   8   0   1 |   b = Aluno Regular
  0  12   3   0   6   0   1 |   c = Transf. Interna Por Reopção de Curso
  0   7   2   0   4   0   1 |   d = Transferência
  0  19   4   0 110   2   0 |   e = Abandono
  0   6   0   0  16   3   0 |   f = Cancelamento
  0   2   0   0   2   0   1 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 43 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Reprovações em Química Geral Experimental entre estudantes que ingressam na faculdade antes do ano de 2010 podem ter influencia na decisão de evadir do curso. Entre 65 estudantes que obtiveram nota final na disciplina variando entre 0 e 3 ocorreram 58 casos de abandono.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 44 – EE – Algoritmo JRip – Experimento 4

```
(GEOPROCESSAMENTO E TOPOGRAFIA <= 4.7) and (ANO_INGRESSO <= 2010) =>
FORMA_EVASÃO=Abandono (18.0/6.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         285               71.25  %
Incorrectly Classified Instances       115               28.75  %
Kappa statistic                          0.557

=== Confusion Matrix ===

  a   b   c   d   e   f   g   <-- classified as
 18   1   0   1   0   0   1 |   a = Formado
  6 151   8   1  12   0   0 |   b = Aluno Regular
  0  10   7   0   5   0   0 |   c = Transf. Interna Por Reopção de Curso
  0   7   2   0   5   0   0 |   d = Transferência
  0  20   2   2  10   4   1 |   e = Abandono
  0   5   0   0  17   3   0 |   f = Cancelamento
  0   5   0   0   0   0   0 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 44 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e utilizando os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Reprovações em cadeiras especificas, como Geoprocessamento e Topografia, também estão relacionadas à evasão. Ocorreram 12 casos de abandono entre os alunos que ingressaram no curso de Engenharia de Energias Renováveis e de Ambiente até 2010 e obtiveram como nota final na disciplina variações entre 0 e 4,7.

A estatística de Kappa demonstra que a concordância dos resultados é moderada e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 45 – EE – Algoritmo PART – Experimento 1

```
ANOS_CURSADOS <= 2 AND
FORMA_INGRESSO = Processo Seletivo - Vestibular AND
CALCULO I <= 7.3 AND
SEXO = F: Abandono (10.66/1.66)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         307               76.75   %
Incorrectly Classified Instances        93               23.25   %
Kappa statistic                          0.6469

=== Confusion Matrix ===

   a   b   c   d   e   f   g   <-- classified as
  13   2   3   0   3   0   0 |   a = Formado
   2 172   0   0   4   0   0 |   b = Aluno Regular
   4   2   5   1   9   1   0 |   c = Transf. Interna Por Reopção de Curso
   3   1   2   2   4   2   0 |   d = Transferência
   4   9   4   2 109   6   1 |   e = Abandono
   0   1   0   1  17   6   0 |   f = Cancelamento
   0   0   2   1   2   0   0 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 45 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “?” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

No cenário descrito no Quadro 45, entra as alunas do sexo feminino que ingressaram através do vestibular e que cursaram até dois anos de faculdade, é demonstrada que entre as estudantes com notas inferiores ou iguais a 7,3 em Calculo I, ocorreram 9 abandonos dentre as 10 alunas pertencentes ao perfil.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 46 – EE – Algoritmo PART – Experimento 2

```
ANO_INGRESSO <= 2011 AND
GESTÃO E PLANEJAMENTO AMBIENTAL = Reprovado por Frequência AND
ALGORITMOS E PROGRAMAÇÃO = Reprovado por Frequência: Abandono (8.22/1.02)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         302               75.5    %
Incorrectly Classified Instances        98               24.5    %
Kappa statistic                          0.6187

=== Confusion Matrix ===

  a   b   c   d   e   f   g   <-- classified as
 16   2   2   0   1   0   0 |   a = Formado
  0 172   0   0   6   0   0 |   b = Aluno Regular
  1   7   1   2  11   0   0 |   c = Transf. Interna Por Reopção de Curso
  1   2   0   0  11   0   0 |   d = Transferência
  7   8   4   0 111   5   0 |   e = Abandono
  1   2   0   0  20   2   0 |   f = Cancelamento
  0   0   3   0   2   0   0 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 46 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “?” e “Aproveitamento” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Entre os estudantes que ingressaram na faculdade até o ano de 2011 e reprovaram por frequência das disciplinas de Gestão e Planejamento Ambiental e Algoritmos e Programação ocorreram 7 abandonos, indicando relações de abandonos com reprovações em mais cadeiras que estão distribuídas no inicio do curso dentro da grade curricular de Engenharia de Energias Renováveis e de Ambiente.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos.

Quadro 47 – EE – Algoritmo PART – Experimento 3

```
ANOS_CURSADOS <= 0 AND
GEOMETRIA ANALITICA <= 2.3 AND
INTRODUÇÃO A ENGENHARIA DE ENERGIA E AMBIENTE <= 0.23: Abandono (33.0/2.0)

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         297               74.25 %
Incorrectly Classified Instances       103               25.75 %
Kappa statistic                          0.6143

=== Confusion Matrix ===

   a   b   c   d   e   f   g   <-- classified as
  21   0   0   0   0   0   0 |   a = Formado
   2 156   6   2   9   2   1 |   b = Aluno Regular
   0   6   6   4   4   2   0 |   c = Transf. Interna Por Reopção de Curso
   0   7   3   0   4   0   0 |   d = Transferência
   0  10   2   4 108   8   3 |   e = Abandono
   0   2   2   0  17   4   0 |   f = Cancelamento
   0   1   1   0   1   0   2 |   g = Transferência Interna
```

Fonte: Próprio Autor

Os resultados apresentados no Quadro 47 foram gerados a partir da mineração de dados de um arquivo Engenharia de Energias Renováveis e de Ambiente, onde foi utilizada a configuração de pré-processamento de dados “0” e “Notas” e sem utilizar os campos destinados para quantidade de vezes que o aluno cursou a disciplina.

Como ultimo experimento realizado no curso de Engenharia de Energias Renováveis e de Ambiente, pode-se analisar mais uma ligação com reprovações em disciplinas básicas e específicas e a evasão. Entre os 33 alunos que cursaram menos de um ano de faculdade e reprovaram em Geometria Analitica e Introdução a Engenharia de Energia e Ambiente, com notas iguais ou inferiores a 2,3 e 0,23, respectivamente, ocorreram 31 casos de abandono.

A estatística de Kappa demonstra que a concordância dos resultados é substancial e a matriz de confusão mostra que os alunos classificados como abandono foram avaliados de maneira correta na grande maioria dos casos, porém a classificação de Transferência Interna por Reopção de Curso apresentou uma taxa alta de erros.

4.6 Análise dos Resultados

Os resultados obtidos para o curso de Engenharia de Computação demonstram que a evasão está ligada a reprovações em disciplinas presentes entre o primeiro e segundo semestres. Com relação às disciplinas específicas do curso, as disciplinas de Introdução a Arquitetura de Computadores, Introdução a Engenharia de Computação e Algoritmos e Programação foram as que demonstraram ligação com abandonos. Em relação a disciplinas básicas do curso, as cadeiras que apresentaram os resultados mais relevantes foram Calculo I, Física I, Laboratório de Física I e Geometria Analitica. O estudo também mostra que, apesar da forte ligação da evasão com o primeiro ano de curso, ocorreram casos de abandono entre alunos que cursaram mais de três anos. Os experimentos também revelaram que aprovações em cadeiras básicas associadas a reprovações em cadeiras específicas levaram alguns alunos a realizarem a reopção de curso dentro da universidade. Com relação a forma de ingresso na faculdade, foi analisado que estudantes que ingressaram como portadores de diploma ou transferência apresentaram tendências à evasão.

O curso de Engenharia de Alimentos também apresenta uma relação entre abandonos e reprovações em disciplinas básicas. Reprovações em Calculo I, Física II, Laboratório de Física II e Química Geral foram apresentadas como prováveis causas de abandonos e cancelamentos entre alunos do curso. Também foi possível analisar que mesmo com a relação com disciplinas iniciais, ocorreram casos de evasão entre alunos com mais de um ano de faculdade. Entre os casos de reopção de curso, foram apresentados casos de alunos que reprovaram em disciplinas específicas do curso de Engenharia de Alimentos, como Química de Alimentos e Físico-Química II.

Os experimentos realizados com os dados relativos aos estudantes de Licenciatura em Física mostraram um cenário um pouco diferente em relação aos cursos de engenharia ofertados na UNIPAMPA Bagé. Foram detectados casos de reopção de curso mesmo quando os alunos obtiveram aprovação em cadeiras de Física I e Laboratório de Física I, que são bases do curso, aliadas a reprovações em cadeiras que são destinadas à área de educação, como Fundamentos da Educação I, História da Educação e Instrumentação para o Ensino de Física I. Esse comportamento pode indicar que alguns estudantes utilizam o curso de Licenciatura

em Física como uma porta de entrada para a faculdade, optando pela reopção de curso posteriormente, devido ao fato de que o ponto de corte em Licenciatura em Física geralmente é mais baixo do que entre os cursos de engenharia. Com relação aos estudantes que abandonaram ou realizaram o cancelamento do curso, ocorrem relações com reprovações em disciplinas presentes nos primeiros semestres do curso. Foram detectadas relações com as disciplinas de Calculo I, Física I, Física II, Laboratório de Física I, Algoritmos e Programação, Química Orgânica e Organização Escolar e Trabalho Docente.

O curso de Engenharia de Produção segue o padrão analisado entre os outros cursos de Engenharia do campus, onde as disciplinas listadas como prováveis causas da evasão estão distribuídas durante os primeiros semestres do curso. Entre as disciplinas básicas, o mau desempenho nas cadeiras de Calculo I, Calculo II, Laboratório de Física I, Produção Acadêmico Cientifica e Economia Industrial apareceram entre as possíveis causas de evasão. Entre as cadeiras especificas, podem ser citadas Sistemas Produtivos I, Sistemas Produtivos II, Engenharia Econômica II, Ergonomia II e Gestão de Qualidade II. Da mesma maneira que foi observado em Engenharia da Computação, durante as analises das possíveis causas da evasão no curso de Engenharia de Produção foi observado que as formas de ingresso como Portador de Diploma, Transferência Voluntária ou Externa e Transferência EX-Officio apresentaram um elevado numero de estudantes evadidos.

Os resultados obtidos durante a mineração de dados do curso de Engenharia Química também seguem o padrão que foi obtido nos outros cursos estudados. Reprovações em disciplinas básicas como Geometria Analitica, Física I, Laboratório de Física I, Probabilidade e Estatística e Algoritmos e Programação podem influenciar na decisão de abandono dos estudos. Entre as disciplinas específicas do curso que foram apresentadas nos estudos, reprovações nas cadeiras de Química Geral Experimental, Introdução a Engenharia Química e Higiene e Segurança no Trabalho também podem influenciar no fenômeno de evasão.

Como ultimo curso analisado, os resultados obtidos na mineração de dados do curso de Engenharia de Energias Renováveis e de Ambiente mostram que o perfil dos estudantes que evadem segue o padrão mostrado nos outros cursos analisados. Reprovações em disciplinas iniciais possuem influência na decisão de abandonar os estudos. Como disciplinas básicas que estiveram presentes nos

experimentos realizados, podem ser citadas Geometria Analitica, Química Geral, Química Geral Experimental, Física I, Calculo I, Fundamentos de Administração e Algoritmos e Programação. Entre as disciplinas especificas Introdução a Engenharia de Energia e Ambiente, Geoprocessamento e Topografia e Gestão e Planejamento Ambiental também aparecem nos experimentos.

De maneira geral, a evasão na UNIPAMPA está fortemente ligada a reprovações em disciplinas iniciais. Disciplinas que são comuns entre todos os cursos estudados, como disciplinas de Calculo, Física e Química possuem influência no fenômeno de evasão. Esse fato pode ocorrer devido a dificuldades de aprendizagem que podem ser trazidas desde antes da vida acadêmica. Cada curso também possui alunos que abandonam em consequência de reprovações em disciplinas específicas. Isso pode ocorrer devido a erros na escolha da profissão, onde o conteúdo abordado não condiz com o que era esperado pelo aluno. Outros fatores como dificuldades financeiras, a universidade estar situada em outra cidade que não seja a cidade-natal ou dificuldades gerais também podem influenciar no processo de decisão de evadir, porém não é possível determinar com base nos dados fornecidos pelo SIE e pelo SiSU.

5 CONSIDERAÇÕES FINAIS

Com base nos resultados apresentados, as metas traçadas para este trabalho foram atingidas. O MineraPampa permite que os arquivos gerados a partir do SIE da UNIPAMPA possam passar pela etapa de pré-processamento dos dados, realizando os processos de limpeza, integração, transformação e redução dos dados. Também possibilita a concatenação das notas obtidas pelos estudantes durante a prova do ENEM e que utilizaram o SiSU como método de entrada na universidade. Adicionalmente, os diferentes modos de personalização do arquivo de saída permitem diferentes cenários de mineração de dados, possibilitando que novos resultados possam ser obtidos.

Com relação aos experimentos realizados, foi possível identificar quais fatores relacionados à vida acadêmica dos estudantes contribuem para que ocorra o fenômeno de evasão dentro da instituição. Os diferentes perfis gerados serão utilizados como apoio pelos grupos de trabalho criados dentro da universidade para tentar realizar o combate à evasão. Soluções, de maneira conjunta, entre os diferentes cursos da UNIPAMPA que contribuam para um melhor desempenho dos alunos, como monitorias ou aulas de reforço, em disciplinas que estão presentes em todos os cursos podem ser uma alternativa.

Finalmente, o *software* que foi desenvolvido fica disponível para que estudantes, professores e funcionários da instituição utilizem como maneira de estudo aos possíveis fatores que causam a evasão. Futuras melhorias também podem ser realizadas, como uma integração entre o MineraPampa e os componentes do Weka responsáveis por realizar a mineração dos dados. A utilização de diferentes fontes de dados também pode contribuir para que resultados mais precisos possam ser obtidos durante os experimentos.

REFERÊNCIAS

ALVES, T. W., ALVES, V. V. **Fatores determinantes da evasão universitária: uma análise a partir dos alunos da UNISINOS.** IV Encontro de Economia Catarinense, 2010.

BALTAR, V. T., OKANO, V. **Análise de Concordância – Kappa.** Laboratório de Epidemiologia e Estatística, Agosto de 2012. Disponível em: http://www.lee.dante.br/pesquisa/kappa/ Acesso em: 20 de fevereiro de 2014.

BORGES, P. **MEC e universidades estudam planos para combater evasão.** IG, 11/02/2012. Disponível em: <http://ultimosegundo.ig.com.br/educacao/mec-e-universidades-estudam-planos-para-combater-evasao/n1597622390779.html> Acesso em: 30 de abril de 2013.

BORIN, J. M. **Desenvolvimento de um Software Utilizando Técnicas de Mineração de Dados para Análise de Evasão na UNIPAMPA.** Salão Internacional de Ensino, Pesquisa e Extensão. Universidade Federal do Pampa, 2013.

CAMARGO, S. da S. **Mineração de Regras de Associação no Problema da Cesta de Compras Aplicada ao Comércio Varejista de Confecção.** Universidade Federal do Rio Grande do Sul, Abril de 2002.

CAMILO, C. O., SILVA, J. C. da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas.** Universidade Federal de Goiás, Agosto de 2009.

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

Falta de engenheiros faz com que profissão esteja em alta no Brasil. **G1**, 11/03/2013. Disponível em: <http://g1.globo.com/jornal-hoje/noticia/2013/03/falta-de-engenheiros-faz-com-que-profissao-esteja-em-alta-no-brasil.html> Acesso em: 16 de abril de 2013.

FILHO, R. L. L. S. e, MOTEJUNAS, P. R., HIPÓLITO, O., LOBO, M. B. de C. M. **A Evasão no Ensino Superior Brasileiro.** Setembro/Dezembro de 2007.

LANOT, A. J. C. **Mineração de Dados Aplicada na Identificação da Propensão à Evasão na Universidade.** Universidade Federal do Pampa, 2012.

MARTINS, A. C., MARQUES, J. M., COSTA, P. D. **Estudo Comparativo de Três Algoritmos de Machine Learning na Classificação de Dados Electrocardiográficos.** Faculdade de Medicina da Universidade do Porto, Março de 2009. Disponível em: http://www.dcc.fc.up.pt/~ines/aulas/0910/MIM/trabs_ano_anterior/noname-1.pdf Acesso em: 20 de fevereiro de 2014.

MORAES, J. O. de, THEÓPHILO, C. R. **EVASÃO NO ENSINO SUPERIOR: Estudo dos fatores causadores da evasão no Curso de Ciências Contábeis da**

Universidade Estadual de Montes Claros – UNIMONTES. 7° Congresso USP de Iniciação Científica em Contabilidade, 2010.

NOGUEIRA, F. **País perde R$ 9 bilhões com evasão no ensino superior, diz pesquisador.** G1, 07/02/2011. Disponível em: <http://g1.globo.com/educacao/noticia/2011/02/pais-perde-r-9-bilhoes-com-evasao-no-ensino-superior-diz-pesquisador.html> Acesso em: 5 de março de 2013.

PESSOA, A. S. A., SILVA, J. D. S., STEPHANY, S., STRAUSS, C., CAETANO, M., FERREIRA, N. J. **Mineração de Dados Meteorológicos Associada a Eventos Severos no Pantanal Sul Matogrossense.** XXXIII Congresso Nacional de Matemática Aplicada e computacional, 2010.

SILVA, M. P. dos S. **Mineração de Dados – Conceitos, Aplicações e Experimentos com Weka.** Universidade do Estado do Rio Grande do Norte, Instituto Nacional de Pesquisas Espaciais, 2004.

VELOSO, T. C. M. A. **Evasão nos Cursos de Graduação da Universidade Federal de Mato Grosso, Campos Universitário de Cuiabá – Um Projeto de Exclusão.** Universidade Federal de Mato Grosso, 2001.