

UNIVERSIDADE FEDERAL FLUMINENSE  
INSTITUTO DO NOROESTE FLUMINENSE DE EDUCAÇÃO SUPERIOR  
LICENCIATURA EM COMPUTAÇÃO

FRANCISCO DE ASSIS SILVA OLIVEIRA

**TÉCNICAS DE MINERAÇÃO DE DADOS APLICADAS À BASE DE DADOS DO  
EXAME NACIONAL DE DESEMPENHO DOS ESTUDANTES**

Santo Antônio de Pádua  
2016

FRANCISCO DE ASSIS SILVA OLIVEIRA

**TÉCNICAS DE MINERAÇÃO DE DADOS APLICADAS À BASE DE DADOS DO  
EXAME NACIONAL DE DESEMPENHO DOS ESTUDANTES**

Trabalho de conclusão de curso  
apresentado ao curso de Licenciatura  
em Computação, como requisito  
parcial para conclusão do curso.

Orientadora:  
Prof.<sup>a</sup> Dr.<sup>a</sup> Geórgia Regina Rodrigues Gomes

Santo Antônio de Pádua  
2016

FRANCISCO DE ASSIS SILVA OLIVEIRA

**TÉCNICAS DE MINERAÇÃO DE DADOS APLICADAS À BASE DE DADOS DO  
EXAME NACIONAL DE DESEMPENHO DOS ESTUDANTES**

Trabalho de conclusão de curso apresentado ao curso de Licenciatura em Computação, como requisito parcial para conclusão do curso. Área de concentração: Computação e Educação.

Aprovada em 22 de Março de 2016.

**BANCA EXAMINADORA**

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Geórgia Regina Rodrigues Gomes (Orientadora) - UFF

---

Prof. Dr. Gustavo Silva Semaan - UFF

---

Prof.<sup>a</sup>. Dr.<sup>a</sup>. – Margarida Pacheco - UFF

Santo Antônio de Pádua  
2016

## RESUMO

Na busca de um sistema educacional superior com qualidade, é preciso conhecer antes, as ocorrências que acometem a atual situação no processo de formação dos futuros profissionais, e traçar planos e estratégias, que busquem as melhores decisões na resolução destes problemas. O objetivo do trabalho é extrair conhecimento da Base de dados do ENADE, aplicando técnicas de Mineração de Dados. Com tais conhecimentos poderá ser realizada uma análise da situação do ensino nas Universidades e Faculdades do Brasil. A partir da análise do desempenho de estudantes no exame, por curso e região, pode-se interpretar a forma como o crescimento qualitativo do Ensino Superior do país, tem se desenvolvido. O software utilizado neste trabalho foi o WEKA, sendo este, um conjunto de algoritmos de aprendizado de máquina para tarefas de Mineração de Dados, desenvolvido em código aberto emitido sob o General Public License, e as tarefas aplicadas foram extração de regras de associação, e árvores de decisão com os algoritmos APRIORI e o J48GRAFT. Após a aplicação dos algoritmos foram extraídas várias regras, as mais relevantes foram à análise de que o conceito Regular, segundo a regra de associação, obteve a maior ocorrência dentre as instituições avaliadas no exame entre os anos de 2010 e 2013, e ainda, pode-se analisar que os melhores desempenhos foram alcançados pelos estados de Minas Gerais e Paraná, pertencentes a regiões do Sudeste e Sul. Junto a isso, com a aplicação do algoritmo foi possível dizer que as regiões, com exceção da região Sudeste e Sul, conseguiram conceitos Ruins no exame realizado, dentre os anos de 2010 e 2013. Nos estados Norte e Nordeste do país, pode-se concluir que há muitas deficiências que são refletidas nos desempenhos de baixo nível no exame, incluindo, algumas instituições públicas. Tendo o real conhecimento da qualidade educacional refletida pelos resultados do exame, tais órgãos responsáveis pelo Ensino Superior no Brasil, podem tirar proveito para construir novos métodos que levem as instituições de Ensino Superior, demonstrarem melhor qualidade, e principalmente, melhor desempenho em avaliações institucionais, em todo território nacional, a partir do conhecimento adquirido neste trabalho.

**Palavras-chaves:** ENADE. Mineração de Dados. Educacional. Superior.

## ABSTRACT

In search of a higher education system with quality, you need to know before the events that affect the current situation in the future professional training process, and make plans and strategies, to seek the best decisions in solving these problems. The objective is to extract knowledge of ENADE Database, using data mining techniques. With such knowledge can be performed an analysis of the situation of education in universities and colleges in Brazil. From the analysis of student performance on the exam, by course and region, can interpret how the qualitative growth of higher education in the country, has developed. The software used in this work was the WEKA, which is a set of machine learning algorithms for data mining tasks, developed in open source issued under the General Public License, and the applied task were extracting association rules, and decision trees with APRIORI and J48GRAFT algorithms. After application of the algorithms were extracted several rules, the most important were the analysis of the Regular concept, the association rule, had the highest occurrence among the institutions evaluated in the examination between the years 2010 and 2013, and also can If we consider that the best performances were achieved by the states of Minas Gerais and Paraná, belonging to regions of the Southeast and South. Next to that, with the application of the algorithm was possible to say that the regions except the Southeast and South, managed bad concepts in examination, from the years 2010 and 2013. In the northern states and the Northeast, it can be concluded that there are many shortcomings that are reflected in the low performance in the examination, including some public institutions. With real knowledge of educational quality reflected by the test results , such agencies responsible for higher education in Brazil , can take advantage of to build new methods that take the higher education institutions , demonstrate better quality, and most importantly, better performance in institutional evaluations , throughout the country , from the knowledge I acquire in this work.

**Keywords:** ENADE. Data Mining. Education. Higher.

## **DEDICATÓRIA**

Aos meus pais, Celiane Silva Oliveira e Ronaldo Silva de Oliveira, pelo zelo, e apoio, durante toda minha trajetória. E também, pela maior herança deixada a mim, a educação. A eles, declaro meu amor eterno, e todo o meu esforço.

## **AGRADECIMENTOS**

A Deus em primeiro lugar, por toda força, capacidade, e bênçãos. E aos meus pais pela maior herança deixada a mim, a educação.

Aos professores, por todo auxílio e amizade, durante toda a minha formação, e também a coordenação do curso e a diretoria da Instituição.

Enfatizo também, um agradecimento aos amigos conquistados durante esta trajetória, que foram partes essenciais, dividindo momentos de descontração, estudos, discussões, experiências e conquistas.

Em especial agradeço ao minha professora e orientadora Geórgia Gomes, que me auxiliou na elaboração deste trabalho, demonstrando paciência e compreensão, sendo tais características de suma importância para meu crescimento acadêmico.

Dê-me, Senhor, agudeza para entender, Capacidade para reter, método e faculdade para Aprender, sutileza para interpretar, graça e Abundância para falar, acerto ao começar, Direção ao progredir e perfeição ao concluir (...).

*São Tomás de Aquino*



## LISTA DE FIGURAS

Figura 1.	Etapas do Processo de DCBD.....	33
Figura 2.	Exemplo de um arquivo Arff.....	41
Figura 3.	Tabela Parcial do Censo ENADE do ano de 2010.....	44
Figura 4.	Base de dados no formato CSV.....	49
Figura 5.	Base de dados no formato ARFF.....	50
Figura 6.	Interface Principal do WEKA.....	51
Figura 7.	Exemplo de Árvore de Decisão.....	53
Figura 8.	Mapa do Território Nacional. Fonte Guia do Estudante (2011).....	57
Figura 9.	Trecho do Modelo gerado no J48GRAFT aplicado Arquivo ENADE 2010. Arff....	58
Figura 10.	Trecho do Modelo gerado no J48GRAFT aplicado Arquivo ENADE 2011. Arff....	63
Figura 11.	Trecho do Modelo gerado no J48GRAFT aplicado Arquivo ENADE 2012. Arff....	69
Figura 12.	Trecho do Modelo gerado no J48GRAFT aplicado Arquivo ENADE 2013. Arff....	73
Figura 13.	Modelo Gráfico do desempenho geral de Instituições Privadas.....	76
Figura 14.	Modelo Gráfico do desempenho geral de Instituições Públicas.....	78

## LISTA DE TABELAS

Tabela 1.	Etapas e Fases do Processo de DCBD.....	34
Tabela 2.	Atributos Existentes nos censos do ENADE.....	45
Tabela 3.	Tratamento do atributo Nota ENADE.....	46
Tabela 4.	Tratamento do Atributo Nota Contínua ENADE.....	47
Tabela 5.	Tratamento do Atributo Nota Padronizada CE.....	48
Tabela 6.	Tratamento do Atributo Nota Padronizada FG.....	48
Tabela 7.	Regras de associação utilizando os atributos Área de Enquadramento (...)	54
Tabela 8.	Regras de associação utilizando os atributos Categ. Administrativa (...)	55
Tabela 9.	Regras de associação utilizando os atributos Área de Enquadramento (...)	60
Tabela 10.	Regras de associação utilizando os atributos Categ. Administrativa (...)	61
Tabela 11.	Grupos ou classificações dos Cursos de Engenharia avaliados.....	64
Tabela 12.	Regras de associação utilizando os atributos Área de Enquadramento (...)	66
Tabela 13.	Regras de associação utilizando os atributos Categ. Administrativa (...)	67
Tabela 14.	Regras de associação utilizando os atributos Área de Enquadramento (...)	71
Tabela 15.	Regras de associação utilizando os atributos Categ. Administrativa (...)	72
Tabela 16.	Dados gerais de desempenho (Privadas).....	76
Tabela 17.	Dados gerais de desempenho (Públicas).....	77

## Sumário

1. INTRODUÇÃO.....	15
1. 2. MOTIVAÇÃO .....	17
1.3. OBJETIVO .....	17
2. FUNDAMENTAÇÃO TEORICA .....	19
2.1. ENSINO SUPERIOR NO BRASIL .....	19
2.1.1 A Qualidade do Ensino Superior no Brasil .....	22
2.1.2. Avaliação Institucional no Brasil .....	23
2.1.2.1 Censo do ENADE .....	26
2.2. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS .....	29
2.2.1. A Importância da Busca do Conhecimento.....	29
2.2.2. Etapas e Fases do DCBD .....	33
2.2.2.1. Pré-Processamento .....	34
2.2.2.2. Processamento de Dados .....	35
2.2.2.2.1. As Tarefas de Mineração de Dados.....	36
2.2.2.2.2. Os Algoritmos de Mineração de Dados .....	37
2.2.2.3. Pós-Processamento.....	39
2.3. SOFTWARE WEKA.....	39
2.3.1. Arquivos ARFF.....	41
3. METODOLOGIA.....	43
3.1. BASE DE DADOS .....	43
Figura 3. Imagem Parcial da Tabela do Censo ENADE do ano de 2010.....	44
3.2. LIMPEZA E TRATAMENTO DOS DADOS .....	44
3.2.1. Tratamento do Atributo: Nota ENADE .....	46
3.2.2. Tratamento do Atributo: Nota Contínua ENADE.....	47
3.2.3. Tratamento do Atributo - Nota Padronizada CE.....	47
3.2.4. Tratamento do Atributo - Nota Padronizada – FG.....	48
3.2.5. Justificativa para a Seleção dos Atributos.....	48
3.3. ADAPTAÇÃO PARA O FORMATO CSV .....	49
Figura 4. Base de Dados no Formato CSV.....	49
Figura 5. Base de Dados no Formato ARFF. ....	50
3.4. MINERANDO A BASE DE DADOS.....	50

Figura 6. Interface Principal do WEKA.....	51
3.4.1 Algoritmo APRIORI.....	52
3.4.2 Algoritmo J48GRAFT .....	52
4. RESULTADOS .....	54
4.1 APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI .....	54
Figura 8. Mapa do Território Nacional. Fonte Guia do Estudante (2011). .....	57
Figura 9. Trecho do Modelo gerado no J48GRAFT aplicado.....	58
Arquivo ENADE 2010. Arff. ....	58
4.3. APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI .....	60
4.4. APLICAÇÕES DA ÁRVORE DE DECISÃO COM ALGORITMO J48GRAFT ..	63
Figura 10. Trecho do Modelo gerado no J48GRAFT aplicado ao .....	63
Arquivo ENADE 2011. Arff. ....	63
4.5. APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI .....	66
4.6. APLICAÇÕES DA ÁRVORE DE DECISÃO COM ALGORITMO J48GRAFT ..	68
Figura 11. Trecho do Modelo gerado no J48GRAFT aplicado ao .....	69
Arquivo ENADE 2012. Arff. ....	69
4.7. APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI .....	70
4.8. APLICAÇÕES DA ÁRVORE DE DECISÃO COM ALGORITMO J48GRAFT ..	73
Figura 12. Trecho do Modelo gerado no J48GRAFT aplicado ao .....	73
Arquivo ENADE 2013. Arff. ....	73
4.9.1. Instituições Privadas.....	76
Figura 13. Modelo Gráfico do desempenho geral de Instituições Privadas. ....	76
4.9.2. Instituições Públicas.....	77
5. CONCLUSÃO.....	79
REFERENCIAS .....	81

## 1. INTRODUÇÃO

Os diversos programas criados para o acesso as Universidades e Faculdades brasileiras, voltados para uma grande massa populacional, sofreram nos últimos anos, grande crescimento. E com isso, nota-se que o Ensino Superior no país ganhou novas características, frente a sua responsabilidade de impacto social, no campo educacional, e profissional. Andrade (2012) define o INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), como o órgão responsável por coletar Dados Educacionais em diversas modalidades, analisá-los, e expor através de pesquisas no campo educacional, para a manutenção e fiscalização do sistema educacional brasileiro, e para isso criou uma série de avaliações que são utilizadas como ferramentas para tal finalidade.

O ENADE trata-se de um exame, sendo um método de avaliação do Sistema Nacional de Avaliação da Educação Superior (SINAES). É aplicado periodicamente aos estudantes de todos os cursos de graduação, durante o primeiro (ingressantes) e último (concluintes) ano do curso. Por sua vez, o ENADE tem como objetivo o acompanhamento do processo de aprendizagem e do desempenho acadêmico dos estudantes em relação aos conteúdos programáticos previstos nas diretrizes curriculares do respectivo curso de graduação. Desta forma, serão construídos referenciais que permitam a definição de ações voltadas à melhoria da qualidade dos cursos de graduação por parte de professores, técnicos, dirigentes e autoridades educacionais.

Ainda de acordo com o pensamento de Andrade (2012), o crescimento das matrículas no setor público vem sendo expandido graças aos programas de acesso e democratização do Ensino Superior, porém ainda encontram-se relativamente pequenos ao percentual de acesso as instituições privadas. Para demonstrar esse crescimento do Ensino Superior no Brasil, às matrículas públicas ampliaram-se com a rede estadual e federal, sendo que os municípios partipicantes do programa de interiorização das Universidades mostram que, tais instituições tem atendido integralmente a grande demanda de suas populações concentradas em diversas localidades.

Mendes (2007), diz que a expansão da rede pública sofre com a falta de recursos, e faz com que essa ampliação ocorra de modo precário e fragilizando as instituições de maior prestígio científico e acadêmico. Desta forma a qualidade do ensino caiu, provocando a falta de maior ampliação nas regiões onde ainda precisa-se do acesso

acadêmico, causando deficiência no setor público de Ensino Superior. Isso reflete muitas vezes nos dados coletados por meio do exame de avaliação do desempenho dos cursos oferecidos pelas instituições de Ensino Superior.

Pesquisas relacionadas ao crescimento das Universidades no Brasil vêm acontecendo desde a década de 70. Para Braga (2010), no período anterior, as análises e levantamentos de informações em educação eram promovidos de forma mais sistemática, em centros de coleta e processamento de dados ligados somente ao Instituto Nacional de Estudos Pedagógicos – INEP, sendo este vinculado ao MEC. Contudo com a expansão dos programas criados pelo governo para o acesso as Universidades, após a reforma universitária em 1968, a produção e estudos relacionados à ampliação das Universidades no país, aumentou em escala exorbitante.

O objetivo deste trabalho é a aplicação de Mineração de Dados, em uma base de dados construída dos censos divulgados pelo INEP resultantes do exame ENADE, dentre os anos de 2010 a 2013, identificando o recente estado qualitativo do Ensino Superior público e privado no Brasil, e ainda, apresentar qual a qualidade do ensino administrados pelas instituições de Ensino Superior, baseada em seus desempenhos em cada exame realizado.

Segundo Dias (2012), desde o surgimento e expansão dos sistemas computacionais, um dos seus principais campos de aperfeiçoamento tem sido a armazenagem de informações diversas, atribuídas e gravadas como dados. É com a necessidade que surgiu ao longo do tempo de administrar de forma aperfeiçoada essas informações, em que a Mineração de Dados foi elaborada, pelo fato dos sistemas tradicionais de exploração de grandes bases de dados não serem mais eficientes para tal tarefa. A Mineração de Dados é umas das ferramentas tecnológicas mais promissoras e avançadas nos dias atuais. Esse sucesso deve-se ao fato de dezenas, centenas ou até milhares de dados serem processados, organizados e padronizados por este método, tornando a busca por informações mais rápida e confiável.

Com tais resultados analisados a partir deste processo e estudo dos mesmos, alcançaremos resultados ou novos conhecimentos em relação ao crescimento do Ensino Superior no Brasil, tanto em ampliação de acesso, quanto ao crescimento qualitativo de ensino.

## 1.2. MOTIVAÇÃO

Realizar um trabalho que esteja relacionado à tecnologia de Mineração de Dados, a fim de expor ainda mais as qualidades dessa tecnologia, que esta em processo de desenvolvimento, e tem seu uso intensificado continuamente. Junto a isso, a concretização de um trabalho que esteja disposto a levar o leitor, a entender o crescimento qualitativo no sistema de educação superior, nos últimos anos. Desta forma, estabelece-se o uso das ferramentas de Mineração de Dados, junto do levantamento de dados do Ensino Superior brasileiro, como um novo estudo sobre a condição da Educação Superior em todo país, através de dados extraídos de avaliações aplicadas a instituições de Ensino Superior, periodicamente.

A motivação, portanto, para este trabalho, se dá pelo fato de obter novos conhecimentos, por meio de técnicas envolvendo a Mineração de Dados aplicada numa base de dados construída a partir de informações já divulgadas, afim de, através da análise dos algoritmos, extrair sobre o que realmente tem ocorrido com o Ensino Superior brasileiro.

## 1.3. OBJETIVO

O objetivo geral deste trabalho é extrair conhecimento da Base de dados do ENADE, aplicando técnicas de Mineração de Dados. Com tais conhecimentos poderá ser realizada uma análise da situação do Ensino Superior no Brasil. Os objetivos específicos deste trabalho consistem em:

- Demonstrar os conceitos teóricos ligados ao processo de descoberta de conhecimento em bases de dados;
- Aplicar tarefas de Mineração de Dados como forma de avaliar a qualidade do ensino nas Instituições de Ensino Superior;

- Analisar os resultados obtidos com a aplicação dessas tarefas de Mineração de Dados, na interpretação de seus modelos, para a extração do novo conhecimento.



## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. ENSINO SUPERIOR NO BRASIL

Em relação à expansão da Educação Superior no Brasil, desde meados de 1810, observa-se que a cada fase de seu processo histórico, pode-se concluir que a Educação Superior no Brasil encontra-se em grande crescimento.

Segundo Trow (2005) apud Andrade (2012), deve-se ver tal ampliação nos dados estatísticos que são gerados e expostos pelo Censo da Educação Superior e sobre a população, construídos pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A transformação histórica da Educação Superior segue três etapas de extrema importância: o sistema de classes, estabelecidos pela sociedade, o sistema de acesso em massa às Universidades e ainda, o processo de democratização da universidade (acesso universal). De acordo com (TROW, 2005, p. 2):

“Dentre as manifestações do crescimento o do sistema de Educação Superior são expressos por três elementos que se tem similaridade entre si, mas que em contrapartida, desencadeiam problemas distintos, tendo influência na taxa de crescimento de matrícula, o tamanho absoluto do sistema e de instituições de educação superior.”

No caso dos índices de crescimento, o que deixa claro o ritmo de expansão, produz uma grande tensão de administração, principalmente sobre a relação de alunos e professores.

Atualmente, o ingresso cada vez mais comum das parcelas da população, que representam as diferentes facetas da sociedade em relação a etnias, status sociais e econômicos, culturais, religiosos e etc., traz a discussão sobre a política de igualdade e equidade de oportunidades educativas antes vista como um segundo plano, estabelecendo uma pressão sobre os órgãos federais de governo e formulação de regras compensatórias. Trow (2005) apud Andrade (2012), também dizia que, somado a esse fato, o aumento de proporções dos grupos etários, tende e muito, a interferir na organização e na relação interna na instituição de ensino, bem como, a estruturação do currículo de cada curso em particular, criando assim, a diversificação e variação das instituições de Ensino Superior.

Ainda de acordo com o raciocínio de Trow apud Andrade (2012), “A Educação Superior chegou a um nível em que, deixou de ser vista como privilégio e direito restrito, de nascimento e classe social, e passa a ser concebida como um direito para todos com certas qualificações”. Dessa forma o Ensino Superior tende a atender a grande demanda da população brasileira que é constituída por classes, que social e economicamente, encontram-se longe de status que deem acesso mais favorável ao crescimento, e desenvolvimento social.

Assim, o ingresso de um grande número de pessoas na educação nas Universidades, tende a fortalecer os movimentos e programas para alterar os mecanismos de acesso e seleção de estudantes, em face de superação da ideia de Educação Superior como privilégio de classe, que cede lugar a ancoragem social de educação como direito. O Ensino Superior no Brasil atualmente é oferecido por Universidades, Centros Universitários, Faculdades, institutos superiores e centros de educação tecnológica. O candidato a essa modalidade de ensino, pode optar por três tipos de formação: bacharelado, licenciatura e formação tecnológica. Os cursos de pós-graduação são divididos entre *latu sensu* (especializações e MBAs) e *strictu sensu* (mestrados e doutorados).

O indivíduo interessado em estudar nas instituições brasileiras de Ensino Superior tem diversas formas de acessá-las. Estas formas podem ser aplicadas pela própria instituição ou por órgãos especializados, como os vestibulares, atualmente cada vez mais instintos. Devido a tal fato, pode-se compreender a atual influência do processo de ingresso a Universidades e Faculdades pelos programas de acesso criados pelo governo, em que os estudantes brasileiros têm a possibilidade de ingresso nas Universidades e demais centros de Educação Superior em todo país.

Evan (2006) explica que o Exame Nacional de Ensino Médio (ENEM), é uma avaliação Seriada no Ensino Médio, que acontece de forma gradual e progressiva, com provas aplicadas ao final de cada série do ensino médio, sendo este cada vez mais usado pelas instituições, sem distinção. No ano de 2015, mais de 1,6 milhão de alunos se inscreveram para o exame, afim de, embarcar no Ensino Superior. Diversas instituições aplicam, ainda, testes, provas e avaliações de conhecimentos voltados à área do curso que o estudante pretende fazer. Algumas instituições, também optam por processos de seleção baseados em entrevistas ou nas informações pessoais e profissionais dos ingressantes na instituição, como grau de escolaridade, cursos, histórico escolar ou experiência e desempenho profissional na área em que se candidatou.

O Estado brasileiro mantém projetos e programas criados pelos órgãos governamentais federais que facilitam o acesso de alunos e professores à Educação Superior e ajudam a melhorar a qualidade de ensino das instituições de Ensino Superior no país.

Como o principal programa criado pelo governo federal, na primeira década do século XXI, pode-se citar o REUNI (Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais) que tem como o principal objetivo, ampliar o acesso e a permanência na Educação Superior pelas diversas classes que tem o acesso as Universidades e a outros institutos de Educação Superior do Brasil. Evan (2006), diz que, o Governo Federal construiu a meta é dobrar o número de alunos nos cursos de graduação em dez anos, a partir de 2008, e permitir o ingresso de 680 mil alunos a mais nos cursos de graduação oferecidos por diversos polos por todo o país.

Ainda, de acordo com Evan (2006), para alcançar a meta, todas as Universidades Federais aderiram ao programa e apresentaram ao ministério público, planos de reestruturação, de acordo com a orientação do programa. As ações preveem, além do aumento de vagas, medidas como a ampliação ou abertura de cursos noturnos para ampliar o acesso dos mais variados públicos de diversas classes, o aumento do número de alunos por professor, evitando o acesso em massa desproporcional à oferta, a redução do custo por aluno, à flexibilização de currículos dos cursos e o combate à evasão ou desistência dos alunos.

No ano de 2015, aconteceu a reunião de diversos reitores de diversas instituições de Ensino Superior do país, para discutir o desenvolvimento científico e tecnológico das instituições de Ensino Superior. De acordo com tal reunião, as Universidades Federais passaram por um processo de reestruturação e expansão, e isso se deve a mudança ocorrida como base, os princípios da democratização e inclusão. A partir daí, vê-se que de fato os programas do governo, como o REUNI, têm estabelecido várias possibilidades de acesso dos jovens ao Ensino Superior, tanto público, quanto privado. Porém, o número de alunos matriculados esperados está abaixo da expectativa especulada mesmo com a criação de tais programas.

A compreensão desses processos de acesso à universidade revelam os obstáculos ao funcionamento adequado do mecanismo educacional existente no Brasil que oferece um suporte para políticas públicas que incidem sobre a melhoria do acesso e da qualidade de

ensino, e a superação de regras decorrentes das diversas formas sociais que se manifestam no interior de todo o sistema educacional do ensino fundamental, médio e superior.

### 2.1.1 A Qualidade do Ensino Superior no Brasil

Desde os primeiros procedimentos estabelecidos no Brasil para a avaliação do desempenho de cursos de instituições de Ensino Superior até então, pode-se dizer que houve uma evolução sistemática na consistência dos indicadores empregados e expostos.

Aqui são apresentados somente alguns atributos do Ensino Superior no Brasil, em relação ao conteúdo de seus cursos, que de certa forma podem servir de parâmetros para alguma avaliação de qualidade. O viés profissionalizante atravessa, de ponta a ponta, a estrutura formal dos diversos cursos de nível superior. Ou seja, para que uma instituição de Ensino Superior brasileiro, tenha o devido reconhecimento qualitativo, deve mostrar por meio da geração de profissionais qualificados para o mercado de trabalho, que é o principal atributo de qualidade do Ensino Superior.

Segundo Cunha (2004), entende-se, neste sentido, que a qualidade é um instrumento ou um mecanismo que tange a melhoria não somente dos serviços e produtos (no caso os profissionais), mas também da oferta, procura e adequação da própria satisfação do consumidor, figura pelo qual se pode atribuir o mercado de trabalho. No Brasil, o Ministério de Educação (MEC) é responsável pela formulação de regras, atributos e políticas com vista à qualidade educacional.

Ainda, Cunha (2004) diz que, todavia, a avaliação se afastou muito do seu uso como fonte de subsídios para a proposição de políticas de impulso da Educação Superior fazendo com que a mesma esteja sempre preparada para as diferenças que a levem a obter uma melhoria significativa. Percebe-se inclusive, que o sistema de avaliação da Educação Superior no Brasil aproxima-se do colapso. Tanto que, as avaliações dos órgãos governamentais, ainda estão sendo reformuladas frequentemente, e o MEC, órgão responsável pelas avaliações das IESs, publicou em dezembro de 2014, novas regras sobre os instrumentos de avaliação institucional no Brasil. Essas questões emergem para a conclusão de que a qualidade educacional deve ser prioridade no quesito qualidade de ensino.

Esta realidade do Ensino Superior no Brasil é uma problemática estudada e debatida nos mais diversos fóruns e instâncias organizacionais interessados no assunto, pois os mesmos acreditam que somente ela, é capaz de orientar os grandes desafios do Ensino Superior brasileiro com resultados satisfatórios.

Segundo Campos (1994), sabe-se, portanto, que a qualidade educacional das IESs assume um papel importante nos status da educação em geral, onde qualidade, além de ser uma pedra máster, torna-se um mecanismo de destaque, competência e concretização da instituição de Ensino Superior.

A qualidade da Educação Superior no país é o tema que mais repercute na sociedade, já que o trabalho das Instituições de Ensino Superior é perceptível e se apresenta não somente na capacitação profissional para a demanda do mercado, na formação acadêmica para o exercício de uma determinada profissão ou ajuste social do indivíduo ao mercado de trabalho, mas concretiza o desenvolvimento tecnológico e científico da sociedade.

### 2.1.2. Avaliação Institucional no Brasil

Pode-se entender sobre avaliação institucional, como uma série de exames em que são aplicados aos diversos centros de ensino em todo país. São pesquisas elaboradas para a coleta de dados em que possam, após uma análise, dizer a situação de determinada instituição, submetida ao método de avaliação.

No Brasil, o INEP, junto ao Ministério da Educação, é o principal órgão de coleta, análise e divulgação de pesquisas, em relação à educação brasileira. E com a expansão das diversas modalidades educacionais no Brasil, em especial a educação superior, construiu-se o ENADE, o meio pelo qual são coletados dados a respeito da educação superior, para o acompanhamento da qualidade dos cursos ofertados pelas diversas instituições avaliadas em todo o país.

Segundo Evan (2006), diz que no ano de 2004, o Ministério da Educação, divulgou a criação do SINAES (Sistema Nacional de Avaliação da Educação Superior) que tem, dentre seus objetivos, a melhoria da qualidade da Educação Superior no Brasil em suas diversas modalidades, sua expansão, e monitoramento da oferta de matrículas, por meio

dos atuais programas de ingresso a Universidades, com a colaboração do Governo Federal. A avaliação institucional é um dos componentes do Sistema Nacional de Avaliação da Educação Superior (SINAES) e está relacionada (INEP, 2010). Dentre os objetivos deste órgão, pode-se citar:

- À melhoria da qualidade da educação superior;
- À orientação da expansão de sua oferta;
- Ao aumento permanente da sua eficácia institucional e efetividade acadêmica e social;
- Promover o crescimento do senso de responsabilidade social, em que as instituições de Ensino Superior possuem, com sua característica de formação de profissionais capacitados e que possam por meio deles contribuir ao âmbito social, sendo esta sua principal missão pública.

O SINAES instituiu o Exame Nacional de Desempenho dos Estudantes (ENADE), passando a integrar o Sistema de avaliação de cursos e instituições em todo país, com a finalidade de acompanhar o processo de aprendizagem dos estudantes de acordo com as Diretrizes Curriculares de cada curso em particular, oferecidos pelas instituições participantes do exame, e ao mesmo tempo, traçar um perfil para a qualidade da Educação Superior ofertada. Perante essas diretrizes, não se pode negar que para a necessidade de acompanhar e avaliar a Educação Superior no país, principalmente diante do crescimento e oferta do setor, se faz necessário, um sistema que avalie o mesmo, expondo suas deficiências que ao longo do tempo, sejam resolvidas.

Segundo Fonseca (2007), tamanha é a importância do exame, e pode-se concluir isso, pelo fato dos resultados da avaliação servirem de referencial para a busca por melhoria e aperfeiçoamento da qualidade nas instituições, tanto em sua expansão em diversas áreas demográficas, e também na ampliação de características qualitativas, além de, proporcionar o acesso às informações necessárias às tomadas de decisões, planejamentos e investimentos na Educação, pelos órgãos responsáveis pela Educação

Superior em todo país, tornando-se assim, a avaliação externa indispensável nesse momento para análise metódica de todas as instituições de educação superior, tanto públicas, tanto privadas.

Segundo Hilu (2011), o exame voltado aos cursos vem sendo aplicado periodicamente, de forma amostral ou censitária, junto aos alunos ingressantes nas Universidades que tenham cumprido de 7 a 22% da carga horária mínima, dos cursos de Graduação Tradicionais ou entre 7 e 25% da carga horária dos cursos superiores de tecnologia, e os concluintes que tenham cumprido pelo menos 80% da carga mínima dos cursos tradicionais ou que tenham concluído pelo menos 75% da carga horária dos cursos tecnológicos, e prováveis concluintes no ano do exame, que segundo a regras do órgão responsável pelo exame, são exigências que servem como atributos que precisam ser cumpridos para a realização do exame na instituição selecionada, referindo-se ao respectivo curso.

A avaliação é aplicada trienalmente pelo INEP, tendo seu resultado expressado por meio de conceitos ordenados de 1 a 5, que são considerados através de outras variáveis, sendo estas, detalhadas a cada censo divulgado, tomando-se por base, padrões estabelecidos por especialistas e órgãos interessados das diversas áreas do conhecimento.

Os resultados do ENADE, a cada censo divulgado, vêm sendo usados por diversos estudos realizados, a fim de, serem utilizados como cálculo de uma infinidade de índices, conceitos e avaliações aplicados pelo Ministério da Educação.

Segundo Lima (2008), há falta de abastecimento de materiais, dispersão do corpo docente da instituição, falta de reconhecimento com a qualidade de ensino na instituição, e diversas outras consequências, que serão com certeza, refletidas na Educação Superior e no seu desempenho, a cada processo de avaliação existente. Como o ENADE tornou-se uma ferramenta para a distribuição dos investimentos das IES (Instituições de Ensino Superior) pelo Governo Federal, com o intuito de se resguardar movimentos, principalmente estudantis, o mesmo tem deflagrado campanhas de conscientização dos alunos e melhorias, na tentativa de demonstrar o que ENADE representa para estudantes, para IES, e para a Educação Superior nacional, como uma importante ação em que se não levada a sério pela comunidade acadêmica, graves consequências aos investimentos serão concretizadas sobre as instituições de Ensino Superior e outros centros relacionados.

Demonstrando a importância do ENADE, e do quanto seus resultados expressos em censos gerais e detalhados vêm influenciando nos resultados e imagem institucional de diversos centros de Ensino Superior em suas diversas áreas, temos a recente Portaria Normativa nº10 de 2 de Julho de 2009 - onde fixam-se os critérios para dispensa de avaliação in-loco nas instituições de Ensino Superior, onde os pedidos de autorização de curso poderão ter a avaliação in loco supridas, dispensado-se da visita dos avaliadores, se o seu Índice Geral de Curso- IGC tiver o seu conceito igual ou superior a 3, após a conclusão do exame. Tais dados tem a finalidade de indicar a qualidade das instituições de Educação Superior em todo país.

#### 2.1.2.1 Censo do ENADE

Com base em estudos e pesquisas bibliográficas realizadas atualmente sobre o assunto, destaca-se uma avaliação apresentando erros e merecedora de reparos em seus critérios, aonde as Universidades, profissionais e alunos são prejudicados, diretamente, por conceitos e rankings estabelecidos em resultados extraídos do ENADE. Com os dados revelados acima se vê que é de extrema importância o exame do ENADE, para todas as instituições brasileiras de Ensino Superior. Porém com o atual descaso em relação ao exame, pela comunidade acadêmica no país, tem causado variados problemas para o desenvolvimento dos diversos cursos oferecidos pelas instituições de Ensino Superior.

Segundo Rothen (2010), para tal insatisfação da comunidade acadêmica para o exame é o fato de em teoria, caso o resultado das avaliações seja baixo, a instituição privada assina um protocolo estipulando termos e prazo para o cumprimento da melhoria dos pontos fracos. No caso da instituição pública, o órgão do poder executivo responsável por sua manutenção, acompanharia o processo de saneamento. Isso é visto por muitos como um problema grave, já que concordam que o ENADE, seja o principal parâmetro de avaliação de qualidade, julgando a instituição, como promissora e digna de investimento.

A cada censo divulgado, Universidades e Faculdades têm sido expostas, como boas ou ruins, inclusive a qualidade de seus cursos, e isso refletem diretamente nos investimentos para a instituição, os índices de formação de novos profissionais, e os índices de matrículas realizados pelos diversos alunos brasileiros, que graças aos



programas governamentais de acesso ao Ensino Superior, tem crescido e ganhado muita força em todo território nacional. A produção científica também fica em má situação, já que os recursos e investimentos são diminuídos pelo governo, que visto pelo mesmo, não produz o suficiente esperado, que condiz com a quantidade de verba liberada para a instituição. Ainda, a imagem de toda a educação das Universidades brasileiras, por sua vez, fica desvalorizada, sendo alvo de muitas às vezes, de posições que são prejudiciais a professores e alunos de diversas instituições de Ensino Superior. Braga (2010), diz que em contrapartida, os órgãos responsáveis pela elaboração e aplicação do exame, precisam ter como objetivo fundamental a melhor forma de avaliar, respeitando as diversidades e especificidades institucionais e regionais, pensando na constituição das condições de seus currículos, formação de seus profissionais, e em seus alunos.

Seja através do ENADE, do CPC, ou de qualquer outro conceito ou índice de avaliação, contando que o processo analítico seja fiel para só assim disponibilizar conceitos justos e dignos de referência de qualidade das instituições, mas também, o embasamento para tomada de decisões dos futuros ingressantes e profissionais no mercado. Dessa forma, a sociedade poderá acompanhar os conceitos divulgados e escolher as melhores IES ou curso que deseja fazer, e o MEC, renovar ou não reconhecimentos. O SINAES precisa ser aplicado na forma como foi concebido, considerando seus mecanismos, princípios e dimensões, pois só assim, ele poderá verificar as IES (Instituições de Educação Superior) que atendem ou não aos requisitos legais, seja ela pública ou privada, não resumindo ao ENADE, e seus resultados divulgados a cada exame.

Às vezes, parece um tanto utópico, mas a busca pela excelência na forma de avaliar, analisar e divulgar a qualidade educacional deve ser estimulado e aperfeiçoado a cada dia, e o exame do ENADE tem por sua vez, essa qualidade, com apoio de especialistas das diversas áreas que são avaliadas e da comunidade acadêmica em geral, observando-se a transparência, e a abertura aos ajustes de procedimentos que se tornam necessários.

O censo ENADE trata-se do conjunto de informações registradas e organizadas em tabelas relacionadas ao resultado dos exames realizados periodicamente. Como dito anteriormente, o ENADE, tornou-se uma poderosa e importante ferramenta de levantamento sobre o crescimento quantitativo e qualitativo de instituições de Ensino

Superior em todo país, públicas e privadas, e que a partir dessa coleta de dados, é analisado minuciosamente cada curso avaliado, periodicamente.

Apesar do ENADE, ser atualmente uma ferramenta de avaliação expressiva da educação aplicadas ao Ensino Superior em todo país, não é necessariamente o único método de levantamento de dados do mesmo. O Censo da Educação Superior coleta, a cada ano, dados importantes para o levantamento geral da situação das Universidades e outros centros e Ensino Superior tais como vagas disponibilizadas, candidaturas às vagas, ingressos de estudantes, matrículas de novos universitários e concluintes dos cursos, tudo isso a fim de traçar uma análise detalhada da situação do nível de Ensino Superior no país.

Ainda sobre tal censo, são coletados, também, outros dados sobre os cursos de graduação e sequenciais – presenciais e a distância – além de dados que tragam informações sobre o corpo docente e técnico-administrativo, financeiros e de infraestrutura, compreendendo bibliotecas, laboratórios e dentre outras instalações diversas da instituição, equipamentos disponíveis ao desenvolvimento, e outros recursos institucionais.

As instituições que não responderem ao censo, que é obrigatória a toda Universidade e Centros de Ensino Superior reconhecidos pelo MEC, ficam impedidas de participar do ENADE (Exame Nacional de Desempenho de Estudantes) e, portanto, seus alunos ficarão impossibilitados de concluir seus cursos, visto que a participação no ENADE é condição indispensável para a emissão do histórico escolar do aluno devidamente matriculado na instituição.

Ainda, segundo Oliveira (2009), o ENADE, mesmo analisado somente enquanto termômetro de desempenho estudantil é extremamente deficitário uma vez que declarar que se pode “acompanhar o processo educacional” - como afirma o governo - por meio de provas durante o desenvolvimento do curso, ignorando as particularidades regionais e culturais, é no mínimo absurdo. Não se concebe uma avaliação como comparação vazia entre cursos, mas sim como busca pela transformação das instituições de Ensino Superior, o que não existe atualmente.

## 2.2. DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS

Desde que a informática e diversas outras ciências aplicadas à tecnologia da informação sofreram uma grande expansão, imensos volumes de dados têm sido sistematicamente coletados e armazenados em variados sistemas e ambientes computacionais. Entretanto, apenas recuperar informação, não propicia todas as vantagens possíveis ao usuário no melhor processamento de seus dados. E quando já não é mais possível à recuperação dos dados, o processo de Mineração de Dados permite que se investiguem nesses dados conhecimento útil.

### 2.2.1. A Importância da Busca do Conhecimento

De acordo com Camilo (2009), assim como um organismo vivo, as empresas, usando-as como exemplo, recebem informações do meio ambiente e também atuam sobre ele, armazenando essas informações em ambientes especializados para preservar e administrar tais dados coletados. Segundo Costa, Coutinho e Magalhães (2005), a forma mais eficaz de se utilizar esses vastos repositórios (Bases de Dados) é tentar descobrir se há algum conhecimento útil escondido neles, ou somente um padrão que possa dizer, mas detalhadamente sobre determinada ocorrência comum no campo em que os dados se referem. Essa já é uma vantagem suficientemente importante para justificar todo o processo, pois permite conhecer o máximo possível da real situação exposta por determinados dados.

Pensando de forma mais simples nesta ideia de descoberta de novos conhecimentos a partir da aplicação de algoritmos e ferramentas afins, pertencentes à Mineração de Dados, em primeiro lugar, é proposto o exemplo de um banco. Neste banco é solicitada uma análise detalhada sobre a real situação das contas de sua centena de clientes. Isso seria possível apenas com um estudo particular, da situação de cada cliente, porém isso custaria uma quantidade enorme de tempo, custo e mão de obra. Agora neste mesmo exemplo, sendo tratado a partir da utilização da Mineração de Dados, seria possível realizar tal tarefa com vantagens que seria exposto a partir de padrões que diriam com precisão, a situação real da conta de cada cliente. A “situação real” seria o que se chama de novo

conhecimento, pois a partir dos resultados, os interessados iriam adquirir o verdadeiro conhecimento do que acontece. Com isso saberiam, se os clientes têm suas contas com movimentação satisfatória? O banco tem sofrido evasão de clientes? O banco lucrou com o número de clientes atuais? E essas e outras variáveis possíveis, ao longo deste ambiente proposto, seriam respondidas com a Mineração de Dados.

Ainda, de acordo com Camilo (2009), ambientes dos mais variados tipos, tem a tendência de sofrer mudanças sejam internas ou externas, devido às interações, que causam a transição de dados e informações a todo o momento, em diversos lugares. E é por isso que se deve sempre procurar ao máximo, expor tais novos conhecimentos, pois eles são o resultado do que realmente se tem feito e concretizado em determinada situação que ocorre. Assim como o exemplo dado acima, é possível citar algumas áreas em que a Mineração de Dados é aplicada, que segundo Camilo (2009), elas são:

- **Marketing:** Ferramentas de Mineração de Dados são aplicadas para descobrir preferências do consumidor e padrões de compra em que são realizadas periodicamente, com o objetivo de realizar marketing direto de produtos e ofertas promocionais aos públicos interessados a determinado produto, de acordo com o perfil do consumidor.
- **Detecção de fraudes:** Muitas fraudes óbvias (tais como, a compensação de cheque por pessoas falecidas) podem ser encontradas sem Mineração de Dados, mas padrões mais sutis de fraude, como essa, podem ser difíceis de serem detectados.
- **Medicina:** Caracterizar comportamento de paciente para prever visitas, identificar terapias médicas de sucesso direcionadas a seus pacientes, para diferentes patologias, e ainda, buscar por padrões de novas doenças, que possam ser prevenidas por pesquisas que desenvolvam diversos medicamentos.
- **Instituições governamentais:** Descoberta de padrões para melhorar as coletas de taxas ou descobrir fraudes que prejudiquem o levantamento de dados nas pesquisas pelos órgãos governamentais.

- **Ciência:** Técnicas de Mineração de Dados podem auxiliar cientistas em seus estudos, por exemplo, encontrar padrões em estruturas moleculares, em física e química, dados genéticos, em citologia e engenharia genética, mudanças globais de clima, em meteorologia, e etc., oferecendo conclusões valiosas rapidamente para suas bases teóricas e aplicáveis.
- **Controle de processos e controle de qualidade:** Auxiliar no planejamento estratégico de linhas de produção, de administradores de grandes empresas, e buscar por padrões de condições físicas na embalagem e armazenamento de produtos.
- **Transporte:** Determinar as escalas de distribuição entre distribuidores, analisar padrões de carga, assunto voltado às empresas especializadas em logísticas.
- **Web:** A busca de padrões que levem ao número de informações sobre determinada área, de diversos assuntos, disponibilizada na rede mundial de computadores.
- **Educação:** A EDM (Mineração de Dados Educacionais) é definida como a área de pesquisa que tem como principal objetivo o desenvolvimento de métodos, ou tomadas de decisões, para explorar conjuntos de dados coletados em ambientes educacionais que são gerados e divulgados, periodicamente.

Como contribuinte, para a busca de conhecimento útil para o aperfeiçoamento de tomada de decisões, é válido, analisarmos a origem de tal método, a Mineração de Dados, e os principais conceitos por traz do desenvolvimento de tal ferramenta. O campo de estudo por traz da tecnologia de Mineração de Dados, a inteligência artificial, que ganha neste trabalho, breve destaque para que o leitor possa conhecer mais essa área que recentemente vem ganhando força dentre os desenvolvedores de sistemas de informação. Para Amo (2003), a construção de um sistema de inteligência artificial envolve uma série de decisões que abrangem vários ramos do conhecimento e áreas especializadas em diversos estudos, tais como a Ciência da Computação, a Pedagogia, a Psicologia, e

obviamente, a área-objeto, ou domínio de conhecimento dos sistemas de inteligência artificial. A Ciência da Computação, e particularmente a Inteligência Artificial, oferece recursos para a representação do conhecimento, modelagem de agentes que serão utilizados nos estudos, heurísticas, e todo o suporte computacional de interface com o usuário; a Pedagogia provê as teorias e modelos de ensino-aprendizagem que tal sistema irá assumir para sua interação com o usuário, e toda a base de anos de prática de ensino com mídias de informação diversas; a Psicologia provê o conteúdo teórico sobre cognição (aprendizagem), emoção (reação do usuário ao sistema) e motivação.

Para a concepção de um ambiente propício à aprendizagem, os aspectos pedagógicos, motivacionais e emocionais, as inteligências artificiais devem estar integradas à arquitetura do sistema, para que este obtenha tal característica de um sistema inteligente. Entretanto, apenas recentemente os aspectos motivacionais e emocionais passaram a ser objetos de interesse dos pesquisadores de sistemas a base de inteligência artificial. Ainda, Amo (2003) diz que, as interferências emocionais no desenvolvimento de um usuário, não são novidades para os profissionais deste campo de estudo. Profissionais de educação, por exemplo, sabe que, estudantes ansiosos, zangados ou deprimidos não aprendem, comprovando que as pessoas, que no caso, os usuários de sistemas computacionais, nesses estados emocionais descritos, não absorvem eficientemente a informação apresentada e disponível, desprezando até mesmo sua qualidade e veracidade.

Existe uma área interessante em que se pode citar como um campo metodológico específico para a análise de Dados Educacionais, chamada de Mineração de Dados Educacionais, em que faz uso da padronização dos dados, para a descoberta de conhecimento relevante para a educação. Baker, Isotani e Carvalho (2011), explicam que a EDM é definida como a área de pesquisa que tem como principal objetivo o desenvolvimento de métodos, ou tomadas de decisões, para explorar conjuntos de dados coletados em ambientes educacionais que são gerados e divulgados, periodicamente. Há muitos métodos utilizados em EDM, que são originalmente da área de Mineração de Dados.

### 2.2.2. Etapas e Fases do DCBD

O DCBD é a descoberta de conhecimento em bases de dados voltado ao processo de tomada de decisão. É o novo conhecimento descoberto e extraído dos dados, por meio das tarefas de Mineração de Dados. Mas detalhadamente, Fayyad (1996) apud Silva (2010), descreve as etapas de DCBD como: a seleção dos dados, o pré-processamento para a extração de dados inconsistentes, transformação para adequá-los ao algoritmo a ser utilizado, a submissão dos dados tratados a tarefas de mineração, e em seguida, a interpretação e validação dos resultados, por meio de modelos gerados, para a identificação do conhecimento. Na Figura 1, observa-se cada fase do processo de DCBD.



Figura 1. Etapas do processo de DCBD. (Adaptado de [Silva, 2010 apud Fayyad et. al., 1996]).

Quintella (2003) descreve três grupos classificatórios, ou etapas do processo de DCBD, como: o **pré-processamento**, a **Mineração de Dados** e o **pós-processamento**. O **pré-processamento** inclui em seu grupo de classificação, todas as fases que prepararam a base de dados, cujos dados selecionados serão fornecidos como entrada para o(s) algoritmo(s) de **Mineração de Dados**, em que inclui as fases de padronização, agrupamento e investigação de informações ocultas nos dados selecionados com entrada

para determinado algoritmo. O **pós-processamento** contempla a depuração e/ou síntese dos padrões descobertos, expondo a solução para o problema. Observa-se na Tabela 1, cada etapa do processo de DCBD.

Tabela 1 – Etapas e Fases do Processo de DCBD.

ETAPAS DO DCBD	FASES
Pré-Processamento	Seleção de Dados
	Limpeza dos Dados
	Transformação de Dados
Processamento	Mineração de Dados
Pós-Processamento	Interpretação de Modelos Gerados
	Conhecimento

O DCBD sendo um processo constituído de etapas classificadas como pré-processamento, processamento dos dados, e pós-processamento, é necessário, da mesma forma, detalhar a cada uma das fases em que compõe tais etapas do processo.

#### 2.2.2.1. Pré-Processamento

Em um processo de DCBD, primeiramente é realizado a etapa de pré-processamento, responsável por coletar e preparar os dados para o processo de Mineração de dados. As fases em que constituem tal etapa podem ser definidas como:

- **Seleção de Dados:** Prevê a coleta e seleção dos dados. É a classificação ou seleção da base de dados em será tratada, considerando o campo de interesse dos dados, e a fonte.



- **Limpeza dos Dados:** Prevê a análise dos dados coletados, verificando a existência de ruídos (erros aleatórios ou variâncias em medições), tratamento de valores ausentes, e identificação de instâncias e atributos em que serão processados, para a obtenção de conhecimento.
- **Transformação dos Dados:** Dedicar-se à incorporação/criação de novos dados a partir dos já existentes. Que possam ser usados com um auxílio na busca pelo novo conhecimento, oculto na base de dados.

#### 2.2.2.2. Processamento de Dados

Esta etapa é definida como a submissão dos dados já tratados as tarefas de mineração de dados, que por sua vez irão gerar modelos e padrões, possibilitando a extração de conhecimento. A Mineração de dados é a fase em que consiste tal etapa do processo de descoberta e conhecimento em uma base de dados.

Pode-se dizer que dentre os objetivos principais da Mineração de Dados, está à função de descobrir relacionamentos entre dados armazenados em determinado ambiente de preservação e administração de dados, e fornecer padrões, em que a partir deles, possa ser feita uma previsão de tendências futuras baseada no passado, ou seja, ocorrências que já sejam comuns e rotineiras em meio aos dados, podendo dizer qual a tendência de evolução ou desaparecimento da mesma.

Segundo Quintella (2003), os resultados obtidos com o processo de Mineração de Dados podem ser usados no gerenciamento de informação de forma segura partindo de um conhecimento gerado sobre determinada ocorrência, solicitações de informações, tomada de decisão, controle de processo e muitas outras aplicações. A Mineração de Dados pode ser aplicada de duas formas: como um processo de verificação, para comprovação de algo, e como um processo de descoberta, na procura de novas rotinas. No processo de verificação, o usuário sugere uma hipótese do que possa estar ocorrendo acerca da relação entre os dados e tenta prová-la aplicando técnicas como análises estatística e multidimensional sobre um banco de dados contendo informações passadas, por exemplo.

No processo de descoberta não é feita nenhuma suposição antecipada. Esse processo usa somente técnicas, tais como descoberta de regras de associação, árvores de decisão, algoritmos genéticos e redes neurais.

Longo (2007) diz que o ser humano continuamente, ao longo da história de sua evolução, toma decisões ou simplesmente chega a determinadas conclusões baseadas no conhecimento que ele acumula ao longo de sua vida, a partir do estudo de seu cotidiano e tarefas de rotina e da experiência. Com essa evolução, chegamos a um tempo em que toda essa carga de dados e conhecimentos é guardada em ambientes de armazenamentos eletrônicos, mas especificamente, banco de dados. Uma grande parte desses dados está armazenada em bases de dados e podem ser facilmente acessadas pelos seus usuários, porém às vezes o acesso, e processamento, tornam-se quase impossível, pela grande vastidão de dados armazenados.

#### 2.2.2.2.1. As Tarefas de Mineração de Dados

Existem várias tarefas de Mineração de Dados para a extração de padrões. Segundo Quintella (2003), é possível defini-los, como:

A **classificação** é o aprendizado supervisionado. Essa tarefa consiste em classificar um item de dado (exemplo ou registro) como pertencente a uma determinada classe, dentre várias classes previamente definidas. Cada classe corresponde a um padrão único de valores dos atributos que são determinados como “chaves” (demais atributos que caracterizam o exemplo). Inicialmente, um conjunto de dados de treinamento é disponibilizado e analisado, e um modelo de classificação é construído, baseado nesses dados. Então, o modelo construído é utilizado para classificar outros dados, chamados dados de teste, os quais não foram contemplados pelo algoritmo durante a fase de treinamento. Para simplificar, o modelo primeiramente gerado, trata-se de uma variável comparativa com a segunda, para que haja sucesso no modelo final.

A **associação** trata-se de uma regra que tem o intuito de identificar associações (relacionamentos comuns dentre os dados). Registros de dados que, de alguma maneira, estão ou devem estar interligados. Sua premissa básica é encontrar elementos que

implicam na presença de outros elementos em uma mesma transação, ou seja, determinado bloco de dados relacionados por situações ali existentes em que serão expostas. Alguns algoritmos que utilizam os conceitos desta tarefa são as regras de associação, e os padrões sequenciais.

**Agrupamento** é uma tarefa que procura traçar/segmentar populações heterogêneas em subgrupos ou segmentos homogêneos, ou seja, fazer o levantamento de pequenos grupos que formam o padrão de um grande grupo detentor destes dados. Mais precisamente, é o processo de formação desses pequenos grupos de objetos dentro de classes de objetos similares é chamado “agrupamento”. Um grupo é uma coleção de objetos de dados que são similares a outro dentro do mesmo grupo.

Outros desenvolvedores de algoritmos mineradores criam ainda, mais duas classes de códigos, em que são definidos como, **Algoritmos de Regressão** que preveem uma ou mais variáveis contínuas, como lucro ou perda, com base nos outros atributos do conjunto de dados. E ainda, **Algoritmos de análise de Sequência** que para quem o aplica, resumem sequencias periódica ou episódios em dados, como um fluxo de caminho da Web, por exemplo.

#### 2.2.2.2.2. Os Algoritmos de Mineração de Dados

Para Namen (2009), os algoritmos de Mineração de Dados são mecanismos que criam um modelo padrão, chamados de solução. Para criar um modelo, um algoritmo primeiro analisa um conjunto de dados e procura padrões e tendências específicos. Galvão (2009) nos dá alguns exemplos de algoritmos de Mineração de Dados, como: K-means, Redes Neurais e Árvores de Decisão. O *K-means* fornece uma classificação de registros, de acordo com os próprios dados existentes na base. Esta classificação é baseada em análise e comparações entre os valores dos dados da base. Ainda têm-se as Redes Neurais, que são técnicas computacionais que geram modelos matemáticos inspirados nas estruturas neurais de organismos inteligentes (o cérebro de mamíferos, por exemplo) e que adquirem e apresentam conhecimento através da experiência. Por fim a Árvore de Decisão, tratando-se de técnicas em que organizam os dados após o processamento, em uma estrutura de árvore.

O algoritmo usa os resultados dessa análise para definir os parâmetros ou atributos chaves, do modelo de mineração. Esses atributos são aplicados pelo conjunto de dados inteiro para extrair padrões acionáveis e estatísticos detalhados. O modelo de mineração que um algoritmo resulta, pode assumir vários formatos, incluindo:

- Um conjunto de diretrizes que descreve como são classificados produtos em uma transação;
- Uma Árvore de Decisão que prevê se um determinado cliente comprará um produto, processo pelo qual muito utilizado por plataformas sociais na internet, tomando esse fenômeno como exemplo;
- Um modelo matemático que prevê as vendas;
- Um conjunto de clusters (grupos), que descreve como os casos (ocorrências) em um conjunto de dados estão relacionados.

Ainda, Namen (2009) diz que a escolha do melhor algoritmo para uma tarefa específica pode ser um desafio. Embora o usuário possa usar algoritmos distintos para executar a mesma tarefa, cada algoritmo produz um resultado diferente, e alguns, podem produzir mais de um tipo de resultado, gerando uma avaliação que exija análise particular de cada variável resolvida na base de dados. O usuário também não tem que usar algoritmos independentemente. Em uma solução de Mineração de Dados, é possível usar alguns algoritmos para explorar dados e, em seguida, usar outros algoritmos para prever um resultado específico com base nesses dados. Pode-se fazer uso de vários algoritmos em uma solução para executar tarefas separadas, como por exemplo, usando um algoritmo de Árvore de Regressão para obter informações de previsão financeira, e um algoritmo com base em regras (tarefas de associação) para melhor análise de cesta básica. Os modelos de mineração, gerados pelos algoritmos, podem prever valores, produzir resumos de dados, e localizar correlações ocultas, sendo essas soluções chamadas de conhecimento.

### 2.2.2.3. Pós-Processamento

Em um processo de DCBD, é necessária a verificação de qualidade do conhecimento descoberto na base de dados tratados pela Mineração de Dados. E para isso faz-se necessário à análise, avaliação e interpretando dos padrões gerados, que por sua vez, auxiliam no de tomada de decisão. Portanto, a etapa de pós-processamento é constituída das fases de **Interpretação de Modelos**, concretizada por um estudo dos modelos gerados pelos algoritmos utilizados na mineração dos dados, e por fim, o **Conhecimento**, sendo o produto final de todo o processo de DCBD.

## 2.3. SOFTWARE WEKA

Para a aplicação do processo de Descoberta de Conhecimento e Bases de Dados e para a aplicação de Mineração de Dados, primeiramente, é preciso todo um processo de produção de arquivos com formatos específicos, onde os dados relevantes para a geração de conhecimento estejam devidamente organizados e dispostos para que os algoritmos comecem seu processo de padronização.

Segundo Baker, Isotani e Carvalho (2011) sabe-se que empresas, estão produzindo bilhões de novos dados, sobre todos os seus usuários, em escala global. Virtualmente todas as organizações, tanto de pesquisas, quanto comerciais, do mundo inteiro, já estão usando a Mineração de Dados, para obter vantagens que levem a cada uma delas ter melhor desempenho no lucro e produção.

O “Waikato Environment for Knowledge Analysis” (WEKA) é um software gratuito, e de código aberto, que pode ser usado para minerar seus próprios dados e transformar o conhecimento sobre seus dados armazenados em informações úteis para aumentar o auxílio de decisões favoráveis. O WEKA é um produto da Universidade de Waikato (Nova Zelândia) e foi utilizado pela primeira vez em sua forma moderna, na década de 90, mais precisamente, em 1997. Ele usa GNU General Public License (GPL). O software foi escrito na linguagem Java, uma das mais populares dentre a comunidade

desenvolvedora de softwares em todo o mundo, e contém uma GUI para interagir com arquivos de dados e produzir resultados visuais. No entanto, a ferramenta fornece uma API Java bastante poderosa e flexível que permite a sua integração a qualquer tipo de sistema Java.

Com uma API (Interface de Programação de Aplicativos) é possível incorporar o WEKA, como qualquer outra biblioteca, a seus próprios aplicativos para fazer coisas como tarefas de Mineração de Dados automatizadas no lado do servidor do usuário da ferramenta. Através de sua interface gráfica (conhecida como Weka Explorer) é possível conduzir processos de mineração de pequenas bases de dados, pertencentes normalmente a usuários pessoais, concretizando a avaliação dos resultados obtidos e a comparação de algoritmos. Além disso, é possível executar tarefas relacionadas ao pré-processamento de dados como, por exemplo, a seleção e a transformação de atributos, para o processamento dos dados.

O Weka se consolidou como a ferramenta de Mineração de Dados mais utilizada em ambiente acadêmico. Além de fins didáticos, o mesmo por sua vez, tem sido alvo de usabilidade em pesquisas científicas. Foi exatamente a sua adequação para este último tipo de aplicação que a tornou popular. Professores responsáveis pelo ensino de Mineração de Dados ou Business Intelligence (inteligência artificial) em Universidades de todo o mundo, costumam utilizar o Weka como instrumento de apoio para o ensino de conceitos básicos sobre Mineração de Dados, e suas tarefas.

É proposto, neste trabalho, o uso dessa ferramenta a favor do Sistema Educacional Brasileiro, graças à sua capacidade de processar dados que estejam dispostos em uma base, e partir deste processamento, gerar novos conhecimentos. Como foi dito, novos conhecimentos, e o processo de implantação dos mesmos, podem influenciar a realidade de tal meio. Sendo assim a Mineração de Dados, pode trazer a possibilidade de tornar possíveis conhecimentos que levem ao encontro de melhores soluções, para aperfeiçoar ainda mais nosso sistema educacional em todo país, assim como também, expor situações adversas atuais no sistema de educação superior.

### 2.3.1. Arquivos ARFF

Um dos maiores problemas persistentes na utilização da aplicação WEKA, é a criação dos arquivos para o minerador. O WEKA utiliza como arquivo padrão para as tarefas de Mineração de Dados, o formato ARFF (Attribute-Relation File Format), porém o minerador pode aceitar também arquivos CSV (Comma-separated values) para realização das tarefas. Entretanto, é preciso informar, e inserir nos arquivos, os atributos de cada instância. Os arquivos ARFF, aceitam basicamente dois tipos de datatypes (tipo de dados) que são Strings (Nominal) e Numéricos. Segundo Baker, Isotani, Carvalho (2011), geralmente o minerador trabalha com Nominal Values (ou valores qualitativos) em tarefas de associação e classificação, e com atributos numéricos em tarefas de agrupamento, contudo isso não é obrigatório, e nem uma regra padrão. Na Figura 2, se encontra um exemplo de toda a estrutura de um arquivo ARFF, em que traz como exemplo a preparação de dados referentes à pesquisa na situação de determinada classe em uma escola, em que irá gerar um resultado, que será interpretado como um novo conhecimento.

```

@relation Escola
@attribute StatusAluno { Passou, Recuperacao, Reprovado }
@attribute NotaGeral numeric
@attribute Aproveitamento numeric
@attribute BomComportamento { true, false }
@attribute ReuniaoPais { yes, no }
@data
%
% 14 Instancias
%
Passou, 85, 85, false, no
Passou, 80, 90, true, no
Recuperacao, 83, 86, false, yes
Reprovado, 70, 96, false, yes
Reprovado, 68, 80, false, yes
Reprovado, 65, 70, true, no
Recuperacao, 64, 65, true, yes
Passou, 72, 95, false, no
Passou, 69, 70, false, yes

```

Figura 2. Exemplo de um arquivo Arff.

A Figura 2 representa a estrutura de um arquivo do tipo ARFF, formato pelo qual se encontram os arquivos a serem utilizados na geração de modelos pelos algoritmos de mineração, através do WEKA. Pode-se perceber pelo exemplo acima, que em primeiro lugar, é declarado à tabela pelo qual esta relacionada aos atributos, em que estes, são

estabelecidos logo abaixo. Em seguida, são declarados os valores, inseridos a cada atributo, de acordo com os valores das instâncias, presente na base de dados.



### 3. METODOLOGIA

O processamento e análise dos dados presentes neste trabalho, serão tratados metodologicamente através das etapas do processo de DCBD. Tal processo consiste na funcionalidade que agrega e organiza os dados, e encontra dentre eles, padrões e relações relevantes. Nas seções a seguir, serão detalhados os processos realizados na extração do conhecimento.

#### 3.1. BASE DE DADOS

A base de dados utilizados neste trabalho, construída com os censos do ENADE dos anos de 2010 a 2013, disponíveis na plataforma online do Instituto INEP, que é a instituição responsável pela fiscalização e levantamento de estudos estatísticos sobre as diversas modalidades de ensino do país. A etapa de seleção e análise da base de dados coletada, em seu estado inicial, foi à responsável por identificar os dados, ou atributos, a serem trabalhados.

Em seguida, foram realizadas as etapas de limpeza, onde foram filtrados, e tratados os dados, e logo após, foram aplicados os algoritmos de associação e classificação, o APRIORI e J48GRAFT, presentes na ferramenta WEKA, na fase de mineração. Serão apresentadas as imagens em que se constituem a base de dados utilizada neste trabalho, partindo da condição de que a cada atributo das tabelas foram preservados. Será apresentado cada passo, de acordo com as diretrizes estabelecidas pelo processo de DCBD.

Para Baker, Isotani e Carvalho (2011), a seleção de dados prevê a coleta e seleção dos dados. É a fase de seleção da base de dados em que os dados serão tratados, considerando os atributos mais relevantes para o estudo. Esta é a principal fonte na coleta destes dados, sendo os mesmos, expostos pela plataforma online do INEP. A base de dados tem seu estado original disponível pela plataforma, em arquivos no formato XLS. Desta forma, a base de dados selecionada se deu por meio de dados organizados em tabelas, totalizando quatro das mesmas, sendo elas responsáveis por conter os resultados do exame.

Categoria Administrativa	Organização Acadêmica	Número de Cursos na Unidade de Cálculo	Número de Concluintes Inscritos	Número de Participantes Concluintes	Média FG_Conc	Média CE_Conc	Nota Enade Concluintes
PÚBLICA	UNIVERSIDADES	1	13	13	58,9231	61,3615	4,8625
PÚBLICA	UNIVERSIDADES	1	62	59	52,9661	53,5932	3,4090
PÚBLICA	UNIVERSIDADES	1	37	37	63,6081	63,8432	3,7822
PÚBLICA	UNIVERSIDADES	1	40	40	54,2750	62,0075	4,4131
PÚBLICA	UNIVERSIDADES	1	0	0			
PÚBLICA	UNIVERSIDADES	1	83	80	51,7625	57,3638	3,8123
PÚBLICA	UNIVERSIDADES	1	0	0			
PÚBLICA	UNIVERSIDADES	1	50	49	58,5714	51,8204	3,7981
PÚBLICA	UNIVERSIDADES	1	72	71	60,4155	58,3577	3,5350
PÚBLICA	UNIVERSIDADES	1	47	47	54,1745	60,2426	3,5439
PÚBLICA	UNIVERSIDADES	1	28	28	59,3036	56,8536	3,3478
PÚBLICA	UNIVERSIDADES	1	38	38	58,8921	63,0684	3,9733
PÚBLICA	UNIVERSIDADES	1	38	38	60,1842	58,4711	3,8406
PÚBLICA	UNIVERSIDADES	1	0	0			
PÚBLICA	UNIVERSIDADES	1	102	98	50,1153	43,8908	4,1032
PÚBLICA	UNIVERSIDADES	1	34	34	47,0000	53,6500	2,9503
PÚBLICA	UNIVERSIDADES	1	0	0			
PÚBLICA	UNIVERSIDADES	1	40	40	39,0375	47,7150	2,0228
PÚBLICA	UNIVERSIDADES	1	0	0			
PÚBLICA	UNIVERSIDADES	1	56	53	50,3962	51,8528	3,0299
PÚBLICA	UNIVERSIDADES	1	35	35	55,9857	63,5857	3,5696

Figura 3. Imagem Parcial da Tabela do Censo ENADE do ano de 2010.

Na Figura 3 é apresentada parte da tabela original do ENADE de 2010. Pode-se perceber que os censos, são estruturados em tabelas, contendo atributos relacionados à instituição, infraestrutura, curso, corpo docente, e desempenho geral. A sessão a seguir, deixará em evidência a eliminação de parte destes atributos, com o intuito de atender da melhor forma o objetivo da pesquisa.

### 3.2. LIMPEZA E TRATAMENTO DOS DADOS

Segundo Galvão (2007), limpeza e tratamento dos dados preveem a análise dos dados coletados, verificando a existência de ruídos (erros aleatórios ou variâncias em medições), tratamento de valores ausentes, e identificação de instancias e atributos, a serem tratados para a obtenção de conhecimento.

Nesta fase, a base de dados adquiriu alterações, para que o resultado final do processo de Mineração de Dados seja o melhor possível, e possa atender a busca pela solução do problema investigado. Foi realizada na base de dados, a eliminação de atributos

em cada tabela, os quais foram julgados irrelevantes para a geração dos resultados úteis, a identificação e eliminação dos erros existentes dentre as instâncias das tabelas, e reestruturação das tabelas. Também foi realizado um tratamento, ou conversão de valores de alguns atributos na base. Nas tarefas de Mineração de Dados utilizadas para este trabalho, foram utilizados valores “alfanuméricos”, pois as tarefas aplicadas apresentam desempenho melhor, com este tipo de dado.

A Tabela 2 apresenta os atributos existentes na base de dados, expondo quais foram eliminados após toda a limpeza, para a obtenção do resultado final e geração de conhecimento. Na coluna (Atributo) da tabela, podem-se conferir todos os atributos existentes na base de dados, enquanto que na coluna (Situação), podem-se observar quais foram os atributos selecionados para a Mineração de Dados.

Tabela 2 – Atributos Existentes nos Censos do ENADE.

<b>ATRIBUTO</b>	<b>SITUAÇÃO</b>
Ano do Exame	Eliminado
Código área	Eliminado
Área de enquadramento (Curso)	Selecionado
Código da Instituição	Eliminado
Nome da Instituição	Eliminado
Sigla da Instituição	Eliminado
Categoria administrativa (publico ou privado)	Selecionado
Org. Acadêmica (classificação da instituição)	Selecionado
Cod. Município	Eliminado
Município do Curso	Eliminado
UF do Curso	Selecionado
Região do Curso	Eliminado
Número de cursos na Unidade	Eliminado
Código do Curso da Unidade	Eliminado
Concluintes inscritos	Eliminado
Concluintes Participantes	Eliminado
Nota Bruta - FG	Eliminado
Nota Padronizada - FG	Selecionado
Nota Bruta - CE	Eliminado
Nota Padronizada - CE	Selecionado

Nota Contínua do ENADE	<b>Selecionado</b>
Conceito ENADE	<b>Selecionado</b>
Nota da Infraestrutura	<b>Eliminado</b>
Nota de Organização Pedagógica	<b>Eliminado</b>

As próximas tabelas mostram como os atributos Nota do ENADE, e Nota Contínua do ENADE, foram tratados para o processo de Mineração da base de dados. É comum na Mineração de Dados, em casos específicos, de acordo com os algoritmos a serem utilizados, trabalhar com valores alfanuméricos ou nominais, ou seja, dados que sejam qualitativos ou não numéricos, pois estes aumentam a probabilidade de resultados mais conclusivos e estáveis ao fim da mineração, e na geração dos modelos.

### 3.2.1. Tratamento do Atributo: Nota ENADE

Na Tabela 3 é apresentado o atributo *NOTA ENADE*, ou seja, o conceito alcançado ao final do exame, por Instituição. Tal nota é alcançada através de um cálculo de média aritmética contendo todos os atributos inseridos na tabela (censo).

Tabela 3 – Tratamento do atributo Nota ENADE.

<b>Nota ENADE - (Valor) na base</b>	<b>Transformado para:</b>
5	Excelente
4	Ótimo
3	Regular
2	Ruim
1	Péssimo

Tais notas, nos Censos ENADE, eram numéricas. Após uma análise em materiais relacionados com a aplicação de tal exame em instituições de Ensino Superior, concluiu-se

que, podem-se definir tais valores numéricos desta forma nominal, associando ao desempenho obtido pelo curso no exame.

### 3.2.2. Tratamento do Atributo: Nota Contínua ENADE

Na Tabela 4 é apresentado o atributo *NOTA CONTÍNUA ENADE* está relacionada ao atributo visto acima, pois o censo do ENADE depende muito do resultado da Nota Contínua, pois o valor do censo ENADE, é decidido de acordo com os valores entre os intervalos que estão expressos na tabela a seguir, sendo a Nota Contínua a nota conclusiva (exata) obtida no exame.

Tabela 4 – Tratamento do Atributo Nota Contínua ENADE.

<b>Intervalo das Notas contínua</b>	<b>Transformado para:</b>
4 - 5	Excelente
3 - 4	Ótimo
2 - 3	Regular
1 - 2	Ruim
0 - 1	Muito Ruim

### 3.2.3. Tratamento do Atributo - Nota Padronizada CE

Na Tabela 5 é apresentado o atributo *NOTA PADRONIZADA – CE* refere-se a Componente Específico (CE), ou seja, atributos ou características relevantes, específicos de determinado curso avaliado no exame. São também classificadas em intervalos para o cálculo do conceito ENADE, de cada curso.

Tabela 5 – Tratamento do Atributo Nota Padronizada CE.

<b>Intervalo das Notas CE</b>	<b>Transformado para:</b>
4 - 5	Excelente
3 - 4	Ótimo
2 - 3	Bom
1 - 2	Regular
0 - 1	Ruim

### 3.2.4. Tratamento do Atributo - Nota Padronizada – FG

Na Tabela 6 é apresentado o atributo *NOTA PADRONIZADA – FG* refere-se à Formação Geral, ou seja, atributos relevantes, específicos em um curso, que influenciem na formação geral oferecida ao aluno de tal curso, tais como: Disciplinas, Bibliografia, etc. São também classificadas em intervalos para o cálculo do conceito ENADE, de cada curso.

Tabela 6 – Tratamento do Atributo Nota Padronizada FG.

<b>Intervalo das Notas FG</b>	<b>Transformado para:</b>
4 - 5	Excelente
3 - 4	Ótimo
2 - 3	Bom
1 - 2	Regular
0 - 1	Ruim

### 3.2.5. Justificativa para a Seleção dos Atributos

A pesquisa, busca pelo real conhecimento do quanto à qualidade da Educação Superior no Brasil tem se concretizado. Pode-se observar que os dados que permaneceram para o processo de Mineração de Dados, são dados que identificam a avaliação geral do índice de desenvolvimento qualitativo das instituições, e ao mesmo tempo, dados que

identificam a situação das instituições, e cursos avaliados de cada região e estado no país, classificando elas de acordo sua organização acadêmica e categoria administrativa. O intuito não é expor as instituições de Ensino Superior de forma negativa, dando informações que podem até mesmo comprometer seu trabalho e contribuição, mas sim, traçar uma relação dos índices alcançados pelas instituições, por região e estado, para que desta forma, seja exposto a atual situação desta modalidade educacional no país nos últimos anos, segundo os resultados do ENADE.

### 3.3. ADAPTAÇÃO PARA O FORMATO CSV

A ferramenta selecionada para realizar o trabalho de Mineração de Dados foi o WEKA, que possui uma exigência específica para que o processo seja concretizado. O formato CSV, que antes foi mencionado, junto ao formato ARFF, são os formatos de arquivo pelo qual a aplicação utiliza como requisito para processar suas tarefas de mineração. Nas Figuras 4 e 5, são apresentadas as imagens da base de dados com seus dados adaptados para o formato CSV e formato ARFF respectivamente, sendo o último, o tipo de arquivo que será trabalhado na ferramenta Weka.

```

Categoria Administrativa,Organização Acadêmica,UF,Número de Concluintes In:
FG bruta,Média CE bruta,Média FG padronizada,Média CE padronizada,Nota Com:
PÚBLICA,UNIVERSIDADES,MT,13,13,58.9231,61.3615,49.2644,35.7192,4.1947,5
PÚBLICA,UNIVERSIDADES,MT,62,59,52.9661,53.5932,54.9057,37.1528,3.5304,4
PÚBLICA,UNIVERSIDADES,MT,37,37,63.6081,63.8432,67.0203,35.4635,3.1230,4
PÚBLICA,UNIVERSIDADES,MT,40,40,54.2750,62.0075,42.6742,36.6966,3.8803,4
PÚBLICA,UNIVERSIDADES,MT,0,0,,50.6800,38.4920,,SC
PÚBLICA,UNIVERSIDADES,MT,83,80,51.7625,57.3638,47.9615,37.2769,3.2999,4
PÚBLICA,UNIVERSIDADES,MT,0,0,,51.2909,38.6291,,SC
PÚBLICA,UNIVERSIDADES,MT,50,49,58.5714,51.8204,56.8452,41.2333,3.3367,4
PÚBLICA,UNIVERSIDADES,MT,72,71,60.4155,58.3577,51.7761,40.0791,3.4718,4
PÚBLICA,UNIVERSIDADES,MT,47,47,54.1745,60.2426,53.3400,44.2750,3.3539,4
PÚBLICA,UNIVERSIDADES,MT,28,28,59.3036,56.8536,57.7364,42.8939,3.2146,4
PÚBLICA,UNIVERSIDADES,MT,38,38,58.8921,63.0684,53.8612,43.1490,3.2956,4
PÚBLICA,UNIVERSIDADES,MT,38,38,60.1842,58.4711,54.5526,42.5053,3.6572,4
PÚBLICA,UNIVERSIDADES,MT,0,0,,37.8333,34.2333,,SC
PÚBLICA,UNIVERSIDADES,MT,102,98,50.1153,43.8908,62.4735,34.9412,3.2407,4
PÚBLICA,UNIVERSIDADES,MT,34,34,47.0000,53.6500,39.0977,32.1126,3.0474,4
PÚBLICA,UNIVERSIDADES,MT,0,0,,54.4688,42.2562,,SC
PÚBLICA,UNIVERSIDADES,MT,40,40,39.0375,47.7150,47.0833,35.4000,1.9665,3
PÚBLICA,UNIVERSIDADES,MT,0,0,,51.1667,36.0944,,SC
PÚBLICA,UNIVERSIDADES,DF,56,53,50.3962,51.8528,56.1000,36.7933,3.5531,4
PÚBLICA,UNIVERSIDADES,DF,35,35,55.9857,63.5857,54.1857,39.7257,3.5974,4

```

Figura 4. Base de Dados no Formato CSV.

Na Figura 4 é representada a estrutura de umas das tabelas pertencentes à base de dados selecionados para tal análise de Mineração de Dados. No formato CSV, é possível

ser gerado por softwares especializados em planilhas, que neste estudo, foi utilizado o Office Excel, através de sua função de conversão de arquivos.

```

relation ENADE_2010@attribute 'Categoria Administrativa' {PÚBLICA,PRIVADA}
{UNIVERSIDADES,FACULDADES,'CENTROS UNIVERSITÁRIOS'}@attribute UF
{MT,DF,SE,AM,PI,MG,SP,PR,PE,RS,RJ,BA,CE,AL,PA,SC,GO,RN,ES,PB,MS,RO,TO,MA,AC
Inscritos' numeric@attribute 'Número de Participantes Concluintes' numeric@
'Média CE bruta' numeric@attribute 'Média FG_padronizada' numeric@attribute
'continua' numeric@attribute 'Conceito Enade' {5,0,4,0,SC,3,0,2,0,1,0}@data
PÚBLICA,UNIVERSIDADES,MT,13,13,58.9231,61.3615,49.2644,35.7192,4.1947,5.0
PÚBLICA,UNIVERSIDADES,MT,62,59,52.9661,53.5932,54.9057,37.1528,3.5304,4.0
PÚBLICA,UNIVERSIDADES,MT,37,37,63.6081,63.8432,67.0203,35.4635,3.123,4.0
PÚBLICA,UNIVERSIDADES,MT,40,40,54.275,62.0075,42.6742,36.6966,3.8803,4.0
PÚBLICA,UNIVERSIDADES,MT,0,0,?,?,50.68,38.492,?,SC
PÚBLICA,UNIVERSIDADES,MT,83,80,51.7625,57.3638,47.9615,37.2769,3.2999,4.0
PÚBLICA,UNIVERSIDADES,MT,0,0,?,?,51.2909,38.6291,?,SC
PÚBLICA,UNIVERSIDADES,MT,50,49,58.5714,51.8204,56.8452,41.2333,3.3367,4.0
PÚBLICA,UNIVERSIDADES,MT,72,71,60.4155,58.3577,51.7761,40.0791,3.4718,4.0
PÚBLICA,UNIVERSIDADES,MT,47,47,54.1745,60.2426,53.34,44.275,3.3539,4.0
PÚBLICA,UNIVERSIDADES,MT,28,28,59.3036,56.8536,57.7364,42.8939,3.2146,4.0
PÚBLICA,UNIVERSIDADES,MT,38,38,58.8921,63.0684,53.8612,43.149,3.2956,4.0
PÚBLICA,UNIVERSIDADES,MT,38,38,60.1842,58.4711,54.5526,42.5053,3.6572,4.0
PÚBLICA,UNIVERSIDADES,MT,0,0,?,?,37.8333,34.2333,?,SC
PÚBLICA,UNIVERSIDADES,MT,102,98,50.1153,43.8908,62.4735,34.9412,3.2407,4.0
PÚBLICA,UNIVERSIDADES,MT,34,34,47.53.65,39.0977,32.1126,3.0474,4.0
PÚBLICA,UNIVERSIDADES,MT,0,0,?,?,54.4688,42.2562,?,SC
PÚBLICA,UNIVERSIDADES,MT,40,40,39.0375,47.715,47.0833,35.4,1.9665,3.0
PÚBLICA,UNIVERSIDADES,MT,0,0,?,?,51.1667,36.0944,?,SC
PÚBLICA,UNIVERSIDADES,DF,56,53,50.3962,51.8528,56.1,36.7933,3.5531,4.0
PÚBLICA,UNIVERSIDADES,DF,35,35,55.9857,63.5857,54.1857,39.7257,3.5974,4.0
PÚBLICA,UNIVERSIDADES,DF,61,61,65.0328,63.7131,58.5595,32.7202,3.4012,4.0
PÚBLICA,UNIVERSIDADES,DF,54,44,50.7955,58.5341,46.6827,41.7615,3.6767,4.0
PÚBLICA,UNIVERSIDADES,DF,32,27,53.55.8852,53.6849,43.6126,3.5477,4.0
PÚBLICA,UNIVERSIDADES,DF,34,34,9.9176,10.9206,51.4081,42.8646,2.2041,3.0
PÚBLICA,UNIVERSIDADES,DF,13,12,53.7167,69.125,50.3833,40.5067,4.4545,5.0
PÚBLICA,UNIVERSIDADES,DF,0,0,?,?,57.1115,32.9393,?,SC

```

Figura 5. Base de Dados no Formato ARFF.

Na Figura 5 é representada a estrutura de umas das tabelas pertencentes à base de dados selecionados para tal análise de Mineração de Dados. No formato ARFF, é possível ser gerado por ferramentas especializadas em Mineração de Dados, que neste estudo, foi utilizado o WEKA, através de sua função de conversão de arquivos. Na Figura 4, demonstra a aparência da base de dados, depois de aplicarmos a fase de limpeza dos dados, e selecionamos os atributos relevantes para a mineração, que agora, está adaptada para a aplicação das tarefas de Mineração de Dados.

### 3.4. MINERANDO A BASE DE DADOS

Na fase de Mineração de Dados foram aplicadas tarefas de classificação e associação, que preveem uma ou mais variáveis discretas, com base nos outros atributos do conjunto de dados. Ou seja, traçam relações gerais sobre os demais atributos presentes na base de dados. Os algoritmos selecionados para o processo minerador de dados foram o J48GRAFT e o *APRIORI*. Estas foram aplicadas nas quatro tabelas da base, a fim de que se



possa, através da comparação de cada caso tratado em todas as tabelas, apresentar um resultado mais confiável em que tenha um bom fundamento. E compará-los.

Em primeiro lugar será utilizada a tabela referente ao ENADE do ano de 2010. Na Figura 6, é apresentada a tela inicial da ferramenta Weka com os atributos que serão utilizados na Mineração de Dados e à direita a estatística de cada atributo.

The screenshot shows the Weka Explorer interface. The 'Select attributes' tab is active. The 'Current relation' is 'ENADE\_2011' with 7576 instances and 11 attributes. The 'Selected attribute' is 'Categoria Administrativa', which is a nominal attribute with 2 distinct values and 0 missing values. A table shows the distribution: 1 PÚBLICA (2873) and 2 PRIVADA (4703). A bar chart below the table visualizes these counts.

No.	Label	Count
1	PÚBLICA	2873
2	PRIVADA	4703

Figura 6. Interface Principal do WEKA.

Pode-se conferir na Figura 6, toda a estrutura da tela principal (ou inicial) do software utilizado para esta pesquisa. Baker, Isotani e Carvalho (2011) diz que, para seleção dos algoritmos que serão utilizados, é preciso pensar primeiro, de como estão registrados todos os dados, para que a partir de suas aplicações surjam resultados aceitáveis.

### 3.4.1 Algoritmo APRIORI

Segundo Baker, Isotani e Carvalho (2011), este algoritmo foi desenvolvido em 1994, pela equipe de pesquisa do Projeto QUEST da IBM (International Business Machines), a mesma responsável pela criação do software Intelligent Miner. Este algoritmo, ou tarefa de Mineração de Dados, trata-se de um algoritmo em que ao final da análise dos dados existentes em determinada base, gera modelos estruturados em regras, em que interpretamos como associações. Tais regras são estruturadas em condicionais, em que se pode observar da seguinte forma:

A) Grupo A = Estudantes ==> Estudantes = Ensino Médio

B) Grupo B = Professores ==> Professores = Ensino Superior

Nos exemplos acima, A e B, demonstram a forma como são geradas as regras pelo algoritmo APRIORI. O exemplo A diz que: **Se os Estudantes pertencem ao Grupo A, então, tais Estudantes, pertencem ao Ensino Médio.** E da mesma, o exemplo B, diz que: **Se os Professores pertencem ao Grupo B, então estes Professores, pertencem ao Ensino Superior.** Assim, é possível interpretar o que as regras, querem dizer, em cada modelo gerado.

O algoritmo APRIORI, traz também, elementos presentes em cada modelo, que são chamados de Suporte mínimo e Confiança. O primeiro parâmetro é o Suporte, ou seja, o percentual de vezes que determinado conjunto de dados aparece em um maior conjunto ou de dados. Enquanto que, a Confiança, indica o percentual de ocorrência de determinada regra, gerada no modelo.

### 3.4.2 Algoritmo J48GRAFT

Este algoritmo gera uma Árvore de Decisão. Entende-se por árvores de decisão, uma estrutura de classificação em que os dados são organizados e padronizados, em estruturas de dados, classificados de árvores. Martins, Marques e Costa (2009), diz que:

“O modelo de Árvore de Decisão é construído pela análise dos dados de treino e o modelo utilizado para classificar dados ainda não classificados. O J48 gera árvores de decisão, em que cada nó da árvore avalia a existência ou significância de cada atributo individual”.

Desta forma entendemos que: o algoritmo gera determinada classificação, em que através deles é possível a identificação de atributos dentro da base de dados. A Figura 6 detalha a estrutura de uma árvore. Nela, podem-se identificar os nós ou elementos, em que constitui a árvore, e junto a isso, a relação, e também a classificação, em que cada um recebe pelo modelo.

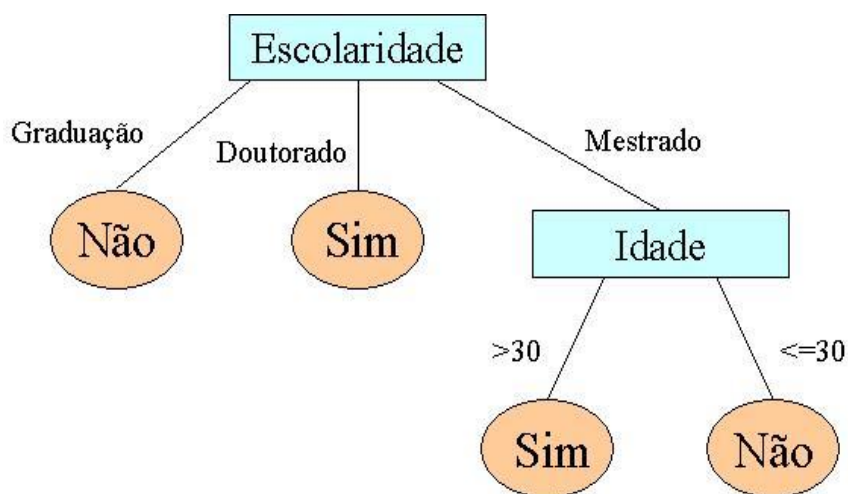


Figura 7. Exemplo de Árvore de Decisão.

## 4. RESULTADOS

Neste capítulo serão apresentados os resultados encontrados na mineração da base de dados selecionada para a pesquisa. O conjunto de informações que aqui serão expostas tem como finalidade, identificar o processo de desenvolvimento qualitativo das instituições de Ensino Superior no Brasil, segundo seu desempenho no exame do ENADE. Durante o processo de Mineração de Dados, foram utilizados algoritmos, de classificação e associação, e isso levou a melhor análise em diversos aspectos da pesquisa, para a descrição final dos resultados, a partir da análise dos modelos descritos.

### 4.1 APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI

Para a fase inicial desta pesquisa, o algoritmo APRIORI, foi utilizado para gerar os principais padrões entre os dados que demonstrem o desempenho das Instituições de Ensino Superior no exame do ENADE. Neste sentido, as Tabelas 7 e 8, destacam os principais padrões encontrados pelo algoritmo aplicado no arquivo: CENSO\_ENADE 2010. Arff e CENSO\_ENADE 2010-2. Arff, representando ocorrências mais relevantes na avaliação realizada no ano de 2010. Esta aplicação serviu para processar informações do desempenho geral de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelas mesmas.

Tabela 7 – Regras de associação utilizando os atributos “Área de Enquadramento, Categ. Administrativa, Org. Acadêmica, Curso, Conceito ENADE, Nota Contínua Conceito ENADE”.

Suporte mínimo	0.2
Confiança Mínima de Padronização	0.9
Nº de ciclos	16
<b>Regras Geradas / CENSO_ENADE 2010. Arff</b>	<b>Confiança</b>
1. Nota Contínua ENADE=bom 1489 ==> Conceito ENADE=Regular 1489.	1.0
2. Conceito ENADE=Regular 1489 ==> Nota Contínua ENADE=bom 1489.	1.0
3. Categoria Administrativa=PRIVADA Nota Contínua ENADE=bom 1269 ==> Conceito ENADE=Regular 1269.	1.0
4. Categoria Administrativa=PRIVADA Conceito ENADE=Regular 1269 ==> Nota Contínua ENADE=bom 1269	1.0

5. Nota Contínua ENADE=Não-Conclusivo 1146 ==> Conceito ENADE=Não conclusivo 1146.	1.0
6. Conceito ENADE=Não-conclusivo 1146 ==> Nota Contínua ENADE=Não-Conclusivo 1146.	1.0
7. Categoria Administrativa = PRIVADA Nota Contínua ENADE=Não-Conclusivo 910 ==> Conceito ENADE=Não conclusivo 910.	1.0
8. Categoria Administrativa=PRIVADA Conceito ENADE=Não conclusivo 910 ==> Nota Contínua ENADE=Não-Conclusivo 910.	1.0
9. Organização Acadêmica=FACULDADES 1454 ==> Categoria Administrativa = PRIVADA 1412.	0.97
10. Categoria Administrativa = PÚBLICA 937 ==> Organização Acadêmica=UNIVERSIDADES 883.	0.94

Tabela 8 – Regras de associação utilizando os atributos “*Categ. Administrativa, Org. Acadêmica, UF, Nota Padronizada – CE, Nota padronizada – FG, Nota Contínua Conceito ENADE*”.

Suporte mínimo	0.35
Confiança Mínima de Padronização	0.9
Nº de ciclos	13
<b>Regras Geradas / CENSO_ENADE 2010 – 2. Arff</b>	<b>Confiança</b>
1. Nota Contínua ENADE=bom 1489 ==> Conceito ENADE=Regular 1489.	1.0
2. Conceito ENADE=Regular 1489 ==> Nota Contínua ENADE=bom 1489.	1.0
3. Nota Padronizada CE=Bom 1489 ==> Conceito ENADE=Regular 1489.	1.0
4. Conceito ENADE=Regular 1489 ==> Nota Padronizada CE=Bom 1489.	1.0
5. Nota Padronizada FG=Bom 1489 ==> Conceito ENADE=Regular 1489.	1.0
6. Conceito ENADE=Regular 1489 ==> Nota Padronizada FG=Bom 1489.	1.0
7. Nota Padronizada CE=Bom 1489 ==> Nota Contínua ENADE=bom 1489.	1.0
8. Nota Contínua ENADE=bom 1489 ==> Nota Padronizada CE=Bom 1489.	1.0
9. Nota Padronizada FG=Bom 1489 ==> Nota Contínua ENADE=bom 1489.	1.0
10. Nota Contínua ENADE=bom 1489 ==> Nota Padronizada FG=Bom 1489.	1.0

Dentre todas as regras geradas pelo algoritmo de mineração, na tabela 7, as regras de número 1, 2, 3, 4, 7, 8, 9 e 10, e na tabela 8, as regras de número 3, 5, 6, 7 e 9 são classificadas as mais relevantes. De acordo com as regras apresentadas, no ano de 2010, grande parte das instituições avaliadas pelo exame mostrou desempenho Regular, em sua

nota final no ENADE, enquanto que em relação às notas contínuas ENADE, receberam classificação como Bom, ou seja, tendo seu resultado contido no intervalo entre Bom e Regular. Tal ocorrência pode-se dizer que, se trata da maior dentre as demais, geradas pela análise dos dados. Assim como tal situação, em terceiro maior número de ocorrências, segundo os padrões gerados, houve grande quantidade de instituições, que de acordo com as regras 5 e 6 da Tabela 6, tiveram seus resultados Não Conclusivos no exame de 2010, tendo sua situação comprometida na conclusão final do exame.

De acordo com as regras 3 e 4 da Tabela 7, as instituições privadas, em 2010, representaram uma grande parcela de instituições participantes do exame, e conseguiram desempenho classificado como Bom, em suas notas contínuas, e obtiveram conceito ENADE classificado como Regular.

A categoria administrativa privada foi a que obteve maior ocorrência em meio às instituições participantes do exame no ano de 2010, tiveram seus resultados não conclusivos no exame, tendo sua situação comprometida na conclusão final do exame, e isto é conferido na regra 7, da Tabela 7.

Segundo as regras 3 e 4 da Tabela 8, o conceito Regular é destacado como a maior ocorrência dentre os resultados finais das instituições avaliadas no ano de 2010, sendo que estas, obtiveram resultados satisfatórios em relação às notas padronizadas CE, ou componentes específicos, classificados com Bom. Sendo que em relação às notas padronizadas FG, em que traz resultados relacionados à formação geral oferecida nos cursos, obteve desempenho semelhante, também classificado como Bom.

Para concluir, nas regras 1, 2, 3 e 4 da Tabela 7, os dados apontam grande vantagem do número de unidades privadas, classificadas em Faculdades, avaliadas em 2010 no exame do ENADE, deixando as Universidades Públicas em pequeno número avaliado. Ainda, diz que tanto as instituições privadas e públicas, formam o grande número de instituições que tornam amplo e mais frequente, o conceito Regular, expondo que os desempenhos gerais das instituições, em 2010, ainda se encontravam em estado não satisfatório.

## 4.2. APLICAÇÕES DA ÁRVORE DE DECISÃO COM ALGORITMO J48GRAFT

Em outra análise na base de dados, foram criadas árvores de decisão utilizando o algoritmo J48GRAFT. A seguir são demonstrados os principais resultados gerados, pela técnica de mineração. Esta aplicação serviu para traçar as informações do desempenho geral de cada estado, categoria administrativa, e organização acadêmica, de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelos cursos, pertencentes às mesmas, através de um “mapeamento descritivo” nos dados do ano de 2010. Ou seja, os resultados serão redigidos sob a exposição específica de cada estado e região do país, através do desempenho de cada instituição de Ensino Superior pertencentes aos mesmos. Na Figura 8 é apresentado o mapa do Brasil e suas Regiões.



Figura 8. Mapa do Território Nacional. Fonte Guia do Estudante (2011).

Tal método será utilizado também para as demais tabelas, em que constitui a base de dados. Na Figura 9, é apresentado um trecho do modelo gerado pelo algoritmo J48GRAFT aplicado a tabela de dados do ENADE 2010.

```

J48graft pruned tree
-----

Categoria Administrativa = PÚBLICA: UNIVERSIDADES (937.0/54.0)
Categoria Administrativa = PRIVADA
| UF = MT
| | Conceito Enade = 5.0: FACULDADES (0.0)
| | Conceito Enade = 4.0
| | | Número de Participantes Concluintes <= 113.5
| | | | Número de Concluintes Inscritos <= 116.5
| | | | | Média FG padronizada <= 61.91025
| | | | | Média FG bruta <= 60.04555
| | | | | | Nota Contínua <= 3.77255
| | | | | | | Média CE bruta <= 59.6932: CENTROS UNIVERSITÁRIOS (3.0/1.0)
| | | | | | | Média CE bruta > 59.6932: UNIVERSIDADES (0.0|43.0/6.0)
| | | | | | | Nota Contínua > 3.77255: UNIVERSIDADES (0.0|12.0/1.0)
| | | | | | | Média FG bruta > 60.04555: UNIVERSIDADES (0.0|31.0/3.0)
| | | | | | | Média FG padronizada > 61.91025: UNIVERSIDADES (0.0|7.0)
| | | | | Número de Concluintes Inscritos > 116.5: UNIVERSIDADES (0.0|8.05)
| | | | Número de Participantes Concluintes > 113.5: UNIVERSIDADES (0.0|8.05)
| | | Conceito Enade = SC
| | | | Número de Concluintes Inscritos <= 3
| | | | | Média CE padronizada <= 24.1468
| | | | | Média CE padronizada <= 21.34055: FACULDADES (0.0|14.0)
| | | | | Média CE padronizada > 21.34055
| | | | | | Média FG padronizada <= 40.77805
| | | | | | Média FG padronizada <= 31.2195: FACULDADES (0.0|11.0/2.0)
| | | | | | Média FG padronizada > 31.2195: UNIVERSIDADES (2.0/1.0)
| | | | | | Média FG padronizada > 40.77805: FACULDADES (0.0|13.0/1.0)
| | | | | Média CE padronizada > 24.1468: FACULDADES (24.0/4.0)

```

Figura 9. Trecho do Modelo gerado no J48GRAFT aplicado  
Arquivo ENADE 2010. Arff.

Com Instâncias classificadas corretamente em aproximadamente 70% na geração dos padrões de dados organizados no modelo gerado a partir do algoritmo J48GRAFT, as Universidades Públicas, apresentam padrões semelhantes a instituições Privadas avaliadas no exame de 2010. Dando início da descrição, sob a visão de um mapeamento descritivo, levantando informações quanto ao desempenho das instituições avaliadas no país, os estados do Mato Grosso, Distrito Federal e Sergipe, pertencentes às regiões do Centro Oeste e Nordeste, apresentam padrões semelhantes, onde as instituições classificadas como Faculdades, segundo o atributo organização acadêmica na base, mostra que nestes estados, os conceitos Regular e Ruim ao final do exame, apresentaram índices em quantidade predominante. Já as Universidades Públicas, em pequena parcela, conseguiram nota classificada como Ótima.

Nos estados do Amazonas e Piauí, na região Norte do país, há grande destaque das instituições classificadas como Centros Universitários e Faculdades. Sendo que estes, em grande parte também concluíram o exame, com notas que predominantemente, se classificam como Regulares e Ruins. Sendo que a nota Regular é a mais comum, dentre estes estados. Mas uma vez, as Universidades Públicas, ganham mais notoriedade, por



alcançarem notas classificadas como Ótimo. Os estados de Minas Gerais e São Paulo, na região Sudeste, mostram padrões semelhantes.

Os estados do Paraná, Rio Grande do Sul, da região Sul, Pernambuco, Bahia, Ceará, da região Nordeste, e Rio de Janeiro na Região Sudeste, apresentam padrões onde às Faculdades particulares, mais uma vez, predominam em números de instituições avaliadas, alcançando resultados iguais a Regular e Ruim. Houve casos, de não conclusão do exame, por uma considerável parcela destas instituições. A nota Regular foi a de grande expressividade dentre estes estados, quase que de 100% dos casos. Enquanto que as Universidades Públicas conseguem uma parcela de notas classificadas em Ótimas, dentre estes mesmos estados.

Os estados de Alagoas, Paraíba, Rio Grande do Norte na região Nordeste do país, Santa Catarina, no Sul, e Espírito Santo, na região Sudeste, houve predominância de número de instituições classificadas em Faculdades, sendo que em relação seus desempenhos, houve índices de alcance de conceitos Ótimos, Ruins e Péssimos. Os Centros Universitários de acordo com os dados foram às instituições que conseguiram conceito Regular, com mais ocorrência. O mesmo para as Universidades Públicas, nestes estados tendo seus conceitos classificados em Ótimos, Ruins e Péssimos.

Nos estados do Mato Grosso do Sul, na região Centro Oeste, Rondônia e Tocantins, na região Norte, em primeiro lugar pode-se dizer que as Universidades Públicas, obtiveram notas que chegaram a Ótimo. Sendo que este índice aplica-se a uma pequena parcela destas instituições, sendo que o maior índice de desempenho dentre este grupo, foi Regular. As Faculdades tiveram desempenhos não satisfatórios nestes estados, sendo iguais a Ruim e Péssimo, havendo casos de não conclusão do exame.

Por fim nos estados do Maranhão, Acre, Roraima e Amapá, região Norte do país, apresentam padrões em que mostram que as Faculdades e Centros Universitários vêm ganhando notoriedade negativa dentre as instituições avaliadas no ano de 2010, sendo que o maior número de casos de não conclusão do exame pertence às Faculdades. E para piorar, não houve nenhuma instituição classificada como faculdade, que obtivesse bons resultados no exame, sendo que todas conseguiram apenas notas que alcançassem conceitos Ruins no exame. Enquanto que os Centros Universitários conseguiram notas classificadas como Ótima, por uma pequena parcela destas instituições, sendo que, a nota Regular, aparece com maior número de ocorrência.

Com esses dados, pôde-se identificar a seguinte situação: Nota-se que no ano de 2010, as Faculdades eram com maior predominância nos números de instituições avaliadas pelo exame. Porém todas as regiões, com exceção da região Sudeste, conseguiram notas muito ruins no exame realizado.

Nos estados do Norte e Nordeste do país, pode-se concluir que há muitas deficiências que são refletidas nos desempenhos de baixo nível no exame, incluindo, algumas instituições públicas. Enquanto nos estados da região Sudeste e Sul, principalmente, as Universidades, possuíram resultados satisfatórios.

#### 4.3. APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI

Novamente, foi aplicado o algoritmo APRIORI para gerar padrões entre os dados que demonstre o desempenho das Instituições de Ensino Superior no exame do ENADE com a Nota Contínua. Neste sentido, as Tabelas 9 e 10, destacam os principais padrões encontrados pelo algoritmo nos arquivos, CENSO\_ENADE 2011. Arff e CENSO\_ENADE 2011-2. Arff, representando ocorrências mais relevantes na avaliação realizada no ano de 2011. Esta aplicação serviu para traçar as informações do desempenho geral de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelas mesmas.

Tabela 9 – Regras de associação utilizando os atributos “Área de Enquadramento, Categ. Administrativa, Org. Acadêmica, Curso, Conceito ENADE, Nota Contínua Conceito ENADE”.

Suporte mínimo	0.1
Confiança Mínima de Padronização	0.9
Nº de ciclos	18
<b>Regras Geradas / CENSO_ENADE 2011. Arff</b>	<b>Confiança</b>
1. Categoria Administrativa = PÚBLICA Nota Contínua ENADE= Ótimo 814 ==> Organização Acadêmica=UNIVERSIDADES 788.	0.97
2. Organização Acadêmica=CENTROS UNIVERSITÁRIOS 928 ==> Categoria Administrativa = PRIVADA 892.	0.96
3. Organização Acadêmica=FACULDADES Nota Contínua ENADE=Bom 904 ==> Categoria Administrativa = PRIVADA 854.	0.94
4. Categoria Administrativa = PÚBLICA Conceito ENADE =Regular 840 ==> Organização Acadêmica=UNIVERSIDADES 785.	0.93
5. Categoria Administrativa = PÚBLICA Nota Contínua ENADE=Bom 895 ==> Organização Acadêmica=UNIVERSIDADES 831.	0.93

6. Categoria Administrativa = PÚBLICA 2873 ==> Organização Acadêmica=UNIVERSIDADES 2642.	0.92
7. Organização Acadêmica=FACULDADES 2245 ==> Categoria Administrativa = PRIVADA 2050.	0.91

Tabela 10 – Regras de associação utilizando os atributos “*Categ. Administrativa, Org. Acadêmica, UF, Nota Padronizada – CE, Nota padronizada – FG, Nota Contínua Conceito ENADE*”.

Suporte mínimo	0.3
Confiança Mínima de Padronização	0.9
Nº de ciclos	14
<b>Regras Geradas / CENSO_ENADE 2011 – 2. Arff</b>	<b>Confiança</b>
1. Nota Padronizada FG=Bom 2593 ==> Conceito ENADE =Regular 2593.	1.0
2. Conceito ENADE =Regular 2593 ==> Nota Padronizada FG=Bom 2593.	1.0
3. Nota Padronizada CE=Bom 2593 ==> Conceito ENADE =Regular 2593.	1.0
4. Conceito ENADE =Regular 2593 ==> Nota Padronizada CE=Bom 2593.	1.0
5. Nota Padronizada CE=Bom 2593 ==> Nota Padronizada FG=Bom 2593.	1.0
6. Nota Padronizada FG=Bom 2593 ==> Nota Padronizada CE=Bom 2593.	1.0
7. Nota Padronizada FG=Bom Nota Padronizada CE=Bom 2593 ==> Conceito ENADE =Regular 2593.	1.0
8. Conceito ENADE =Regular Nota Padronizada CE=Bom 2593 ==> Nota Padronizada FG=Bom 2593.	1.0
9. Conceito ENADE =Regular Nota Padronizada FG=Bom 2593 ==> Nota Padronizada CE=Bom 2593.	1.0
10. Nota Padronizada CE=Bom 2593 ==> Conceito ENADE =Regular Nota Padronizada FG=Bom 2593.	1.0

Dentre todas as regras geradas pelo algoritmo de mineração, na Tabela 9, as regras de número 1, 2, 3, e na Tabela 10, as regras de número 1, 3, 5, 6, são classificadas as mais relevantes. Segundo os padrões gerados pelas regras constituídas pelo algoritmo, uma pequena parcela das instituições avaliadas no ano de 2011, são classificadas como Universidades Públicas, obtiveram notas contínuas classificadas como Ótimo, ou seja, essa mesma parcela de instituições alcançou conceito Ótimo, no final do exame.

Na regra 2 da Tabela 9, mostra que o número e instituições participantes do exame de 2011, referente à organização acadêmica, classificadas em Centros Universitários, era superior ao número de instituições privadas, classificadas como Faculdades. Com isso

percebe-se que em 2011, era realmente visível, o crescimento de instituições privadas de Ensino Superior, sendo estes, os componentes com maior participação nos mecanismos de avaliação nesta modalidade de ensino no país. Outra regra, de número 3 da tabela 9, mostra que as instituições classificadas como Faculdades Privadas, alcançaram resultado classificado como Bom, em sua Nota Contínua ENADE.

Nas regras 4, 5 e 6 da Tabela 9, se refere às instituições públicas participantes do exame, mostrando que pequena parcela alcançou conceito ENADE Regular. Sendo que, instituições classificadas como Universidades, estão com presença considerável dentre esse número de instituições. Nesta mesma ocorrência, pode-se dizer também que, as Universidades conquistam um número maior de instituições em que obtiveram conceito Ótimo ao final do exame.

Observando os padrões gerados pelo algoritmo, nas regras 3 e 4 da Tabela 10, pode-se dizer que em relação à nota de formação geral dos cursos, ou Nota Padronizada FG, os cursos das instituições que alcançaram bom resultado neste quesito, em contrapartida, alcançaram desempenho Regular no conceito final do exame. Pode-se ainda dizer que, em relação à Nota Padronizada CE, referente ao componente específico dos cursos avaliados, as instituições que alcançaram desempenho Regular no conceito final do exame, da mesma forma, tiveram resultados satisfatórios no quesito em questão.

Em relação a números de instituições, classificadas por organização acadêmica, pode-se notar o fato de que no ano de 2011, dentre as instituições públicas avaliadas, indicadas nas regras 4, 5 e 6 da Tabela 9, grande parte das unidades classificadas assim, eram Universidades. Com isso conclui-se que em 2011, o número de Universidades Públicas avaliadas no exame apresentava maior número, em relação às instituições privadas. Estas unidades de Ensino Superior, ganha grande força em seu processo de expansão territorial em todo país, acrescentando que, referindo-se à qualidade, muitas alcançaram resultado satisfatório no exame.

#### 4.4. APLICAÇÕES DA ÁRVORE DE DECISÃO COM ALGORITMO J48GRAFT

Em outra análise na base de dados, foi criada a Árvore de Decisão utilizando o algoritmo J48GRAFT. Na Figura 10, é apresentado um trecho do modelo gerado com a aplicação desta técnica. Este resultado apresenta informações do desempenho geral de cada estado, categoria administrativa, e organização acadêmica, de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelos cursos, pertencentes às mesmas, através de um mapeamento regional nos dados do ano de 2011.

```

Categoria Administrativa = PÚBLICA
| Censo do Enade <= 4.0094
| | Sigla UF = MT: UNIVERSIDADES (3.0)
| | Sigla UF = DF: UNIVERSIDADES (2.0)
| | Sigla UF = SE: UNIVERSIDADES (0.0)
| | Sigla UF = PI: UNIVERSIDADES (1.0)
| | Sigla UF = MG
| | | Número de Concluintes Participantes <= 11: UNIVERSIDADES (5.0/1.0)
| | | Número de Concluintes Participantes > 11: FACULDADES (12.0)
| | Sigla UF = PR
| | | Número de Concluintes Participantes <= 7: UNIVERSIDADES (3.0/1.0)
| | | Número de Concluintes Participantes > 7: FACULDADES (12.0)
| | Sigla UF = PE
| | | Nota de Infraestrutura <= 0.7839: UNIVERSIDADES (2.0)
| | | Nota de Infraestrutura > 0.7839: FACULDADES (33.0)
| | Sigla UF = RS: UNIVERSIDADES (4.0)
| | Sigla UF = RJ: UNIVERSIDADES (26.0/1.0)
| | Sigla UF = SP
| | | Censo do Enade <= 1.5161
| | | | nota Contínua do Enade <= 1.8768
| | | | | Nota de Infraestrutura <= 3.2888: FACULDADES (4.62)
| | | | | Nota de Infraestrutura > 3.2888
| | | | | Nota de Infraestrutura <= 4.3203: UNIVERSIDADES (6.62/1.62)
| | | | | Nota de Infraestrutura > 4.3203: FACULDADES (2.92)
| | | | nota Contínua do Enade > 1.8768: FACULDADES (31.85/1.0)
| | | Censo do Enade > 1.5161
| | | | Média CE bruta <= 35.4033
| | | | Censo do Enade <= 3.4028

```

Figura 10. Trecho do Modelo gerado no J48GRAFT aplicado ao Arquivo ENADE 2011. Arff.

Neste caso, a Árvore de Decisão gerada, com Instâncias classificadas corretamente em 70%, forneceu um padrão comum dentre os estados brasileiros, especificando cada área de enquadramento ou curso, e a organização acadêmica do mesmo, exibindo o desempenho de cada um, no exame realizado em 2011.

Universidades Públicas e Faculdades Particulares, de maneira geral, dos estados do Mato Grosso, Distrito Federal, Sergipe, Piauí, Minas Gerais, Paraná, Paraíba, Pernambuco, Goiás, Bahia, Ceará, Maranhão, Espírito Santo, Alagoas, Mato Grosso do Sul, Roraima, Amapá, Rondônia e Acre, são os estados em que mais tiveram instituições classificadas

como Faculdades, avaliadas no exame. Amazonas e Tocantins tiveram os Centros Universitários, em quantidade predominante de instituições avaliadas. Enquanto que, Santa Catarina, Rio Grande do Norte, Rio Grande do Sul, Rio de Janeiro e São Paulo, as Universidades marcaram grande presença na realização do exame.

Os cursos de Arquitetura e Urbanismo, Artes Visuais, Biologia, Ciências Sociais e Computação, pertencentes predominantemente a instituições classificadas como Universidades, sendo que em sua maior parte, em relação ao desempenho no exame, alcançaram conceito Regular. E ainda, tiveram os índices de desempenho, considerados Ótimo e Excelente, alcançados por poucas instituições, mostrando que uma parcela mínima de Universidades concluiu o exame com êxito. Em relação a estes mesmos cursos, alocados em instituições classificadas como Faculdades, os índices de desempenho classificados como Ruins foi elevado, quase que 50% das instituições obtiveram conceito Ruim e Péssimo.

Os cursos de Engenharia de Computação, Licenciatura em Computação, Sistemas de Informação e Educação Física, provenientes de todas as Organizações acadêmicas, sendo que, o curso de Sistemas de Informação, pertence em grande parte a instituições classificadas como Faculdades. Esses cursos, segundo os dados, obtiveram conceitos Excelentes, e Ótimos nos resultados finais do exame, em instituições classificadas como Universidades. Enquanto que as Faculdades obtiveram em maior número, desempenhos Regulares, sendo que uma boa parte das Faculdades, concluíram o exame com conceito Ruim. Os cursos de Engenharia, classificados em oito grupos, são definidos na Tabela 11:

Tabela 11 – Grupos ou classificações dos Cursos de Engenharia avaliados.

GRUPO	DESCRIÇÃO
A	Engenharia de Infraestrutura (cartográfica, civil, e etc.).
B	Engenharia de Sistemas (computação, redes de comunicação, e etc.).
C	Engenharia de Produção Industrial (automotiva, mecânica, e etc.).
D	Engenharia Química e Biológica (alimento, química, e etc.).
E	Engenharia de Materiais.
F	Engenharia de Produção.
G	Engenharia de Recursos e Energia (minas, petróleo, etc.).
H	Engenharia da Natureza (florestal, agrícola, e etc.).

As que pertencem aos grupos A e B, nas Universidades, obtiveram desempenhos iguais a Ótimo e Ruim, respectivamente. Enquanto que, nas Faculdades, tais cursos concluíram o exame com desempenho Regular.

Nos grupos C e D, em relação às Universidades, obtiveram desempenhos iguais a Regular, Ruim e Péssimo. Enquanto que estes mesmos cursos, nos Centros Universitários, obteve resultado Excelente, no exame.

Nos grupos E, e F, em relação às instituições classificadas como Faculdades, obtiveram conceitos Excelente, Ótimo e Péssimo. Os Centros Universitários tiveram baixo desempenho, alcançando resultado Péssimo, em maior ocorrência. Já as Universidades, predominantemente, conseguiu nota Regular, nestes cursos.

Por fim, nos cursos de engenharia dos grupos G e H, junto dos cursos de Filosofia, Física, Geografia e História, nas instituições classificadas como Centro Universitário, os resultados alcançados foram classificados como Ótimo, e Excelente, enquanto que nas instituições classificadas como Faculdades, obtiveram desempenho Regular. Nas Universidades Públicas, as notas também foram satisfatórias.

Nos cursos de Letras, e Matemática, em relação às Faculdades, mais de 50% destas instituições, obtiveram conceito Regular, porém, parte destes mesmos cursos inseridos nestas unidades, obtiveram resultados Ruins, e Péssimos. Nas Universidades, as notas em pequena parte destas instituições, concluíram o exame, com resultado Excelente.

Nos cursos de Música, Pedagogia, Química, Tecnologia em Alimentos, e Tecnologia em Análise e Desenvolvimento de Sistemas, nas instituições classificadas como Faculdades, tiveram como resultado de seu desempenho no exame, classificações como Ótimo, Ruim e Péssimo. Nas Universidades, os padrões mostram, como maior ocorrência, conceitos Excelentes. Em relação aos cursos de tecnologia em geral (Automoção Industrial, Construção, Gestão, etc.), nos Centros Universitários, os desempenhos tiveram resultado Regular, em sua maior parte. Nas Faculdades, tais cursos obtiveram conceito Ruim. Em especial, os cursos de Tecnologia em Redes de Computadores, e Tecnologia em Saneamento Ambiental, obtiveram conceito Ótimo, em maior ocorrência. Enquanto que as Universidades alcançaram resultados Péssimos em maior ocorrência, cerca de 60%, e conceito Ótimo, por uma parcela mínima, aproximadamente 20% das instituições.

Com esses dados, pode-se dizer que as Instituições, sejam elas Privadas e Públicas, se encontravam em níveis semelhantes, no ano de 2011. Apesar do destaque para algumas

das regiões, em especial o Sudeste do País, muitas Universidades e Faculdades, sofrem como baixo desempenho de seus cursos ofertados e avaliados, causando grande déficit de condições pedagógicas, científica, e de qualidade.

#### 4.5. APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI

Novamente, o algoritmo APRIORI, foi utilizado para criar os principais padrões entre os dados que demonstre o desempenho das Instituições de Ensino Superior no exame do ENADE, relacionando-os com o censo do ENADE, e também, com a Nota Contínua. Neste sentido, as tabelas 12 e 13, destacam os principais padrões encontrados pelo algoritmo nos arquivos CENSO\_ENADE 2012. Arff e CENSO\_ENADE 2012 - 2. Arff, representando ocorrências mais relevantes na avaliação realizada no ano de 2012. Esta aplicação serviu para traçar as informações do desempenho geral de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelas mesmas.

Tabela 12 – Regras de associação utilizando os atributos “Área de Enquadramento, Categ. Administrativa, Org. Acadêmica, Curso, Conceito ENADE, Nota Contínua Conceito ENADE”.

Suporte mínimo	0.25
Confiança Mínima de Padronização	0.9
Nº de ciclos	15
<b>Regras Geradas / CENSO ENADE 2012. Arff</b>	<b>Confiança</b>
1. Nota Contínua do ENADE =Bom 2770 ==> Conceito ENADE =Regular 2770.	1.0
2. Conceito ENADE =Regular 2770 ==> Nota Contínua do ENADE =Bom 2770.	1.0
3. Categoria Administrativa = PRIVADA Nota Contínua do ENADE =Bom 2477 ==> Conceito ENADE =Regular 2477.	1.0
4. Categoria Administrativa = PRIVADA Conceito ENADE =Regular 2477 ==> Nota Contínua do ENADE =Bom 2477.	1.0
5. Nota Contínua do ENADE =Regular 1719 ==> Conceito ENADE =Ruim 1719.	1.0
6. Conceito ENADE =Ruim 1719 ==> Nota Contínua do ENADE =Regular 1719.	1.0
7. Categoria Administrativa = PRIVADA Nota Contínua do ENADE =Regular 1594 ==> Conceito ENADE =Ruim 1594.	1.0
8. Categoria Administrativa = PRIVADA Conceito ENADE =Ruim 1594 ==> Nota Contínua do ENADE =Regular 1594.	1.0
9. Organização Acadêmica=FACULDADES 3370 ==> Categoria Administrativo PRIVADA = PRIVADA 3262.	0.97



10. Conceito ENADE =Ruim 1719 ==> Categoria Administrativa =PRIVADA 1594.	0.93
---	------

Tabela 13 – Regras de associação utilizando os atributos “*Categ. Administrativa, Org. Acadêmica, UF, Nota Padronizada – CE, Nota padronizada – FG, Nota Contínua Conceito ENADE*”.

Suporte mínimo	0.4
Confiança Mínima de Padronização	0.9
Nº de ciclos	12
<b>Regras Geradas / CENSO_ENADE 2012 – 2. Arff</b>	<b>Confiança</b>
1. Nota Contínua do ENADE =Bom 2770 ==> Conceito ENADE =Regular 2770.	1.0
2. Conceito ENADE =Regular 2770 ==> Nota Contínua do ENADE =Bom 2770.	1.0
3. Nota Padronizada FG=Bom 2770 ==> Conceito ENADE =Regular 2770.	1.0
4. Conceito ENADE =Regular 2770 ==> Nota Padronizada FG=Bom 2770.	1.0
5. Nota Padronizada CE=Bom 2770 ==> Conceito ENADE =Regular 2770.	1.0
6. Conceito ENADE =Regular 2770 ==> Nota Padronizada CE=Bom 2770.	1.0
7. Nota Padronizada FG=Bom 2770 ==> Nota Contínua do ENADE =Bom 2770.	1.0
8. Nota Contínua do ENADE =Bom 2770 ==> Nota Padronizada FG=Bom 2770.	1.0
9. Nota Padronizada CE=Bom 2770 ==> Nota Contínua do ENADE =Bom 2770.	1.0
10. Nota Contínua do ENADE =Bom 2770 ==> Nota Padronizada CE=Bom 2770.	1.0

Dentre as regras geradas, as regras de número 3, 4, 7, 8 e 9 na tabela 12, e as regras de número 1, 3, 5, 6 na Tabela 13, são mais relevantes. Nas regras de número 3, 4, 7, 8 e 9 na tabela 11, segundo os padrões gerados pelos dados, no ano de 2012, as notas contínuas do exame ENADE, classificadas como Boas tiveram maior ocorrência em meio as Instituições avaliadas no ano de 2012, assim como, o conceito Regular. As mesmas regras dizem que, em relação a instituições classificadas como unidades privadas, as notas contínuas do exame ENADE, foram classificadas como Boa, porém, seu conceito ENADE, foi igual a Regular. Junto a isso, dentre estas mesmas instituições privadas, que voltam a ser o grupo com maior número de unidades avaliadas em 2012, totalizando 3.262 unidades, demonstra desempenhos, Ruim e Péssimo, tomados como resultados abaixo da média.

Nas regras de número 1, 3, 5, 6 na Tabela 13, os dados vêm ainda apresentar padrões em que dizem que as instituições que concluíram o exame, com resultados classificados como Bom, em seus requisitos referentes à formação geral, ou Nota Padronizada FG, e também, em relação a sua nota de componentes específicos no curso, ou Nota Padronizada CE, tornaram-se principal fator que levou instituições Públicas e Privadas, em grande número, a concluírem o exame com conceito classificado como Regular. Com isso pode-se dizer que dentre instituições Públicas e Privadas, o desempenho foi semelhante, em conceitos que foram classificados como Regular e Ruim, tendo tais conceitos, os mais comuns dentre as instituições.

#### 4.6. APLICAÇÕES DA ÁRVORE DE DECISÃO COM ALGORITMO J48GRAFT

Em outra análise na base de dados, foi criada a Árvore de Decisão utilizando o algoritmo J48GRAFT. A seguir são demonstrados os principais resultados provenientes, desta técnica de mineração aplicada. Esta aplicação serviu para traçar as informações do desempenho geral de cada estado, categoria administrativa, e organização acadêmica, de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelos cursos, pertencentes às mesmas, através de um mapeamento regional nos dados do ano de 2012.

```

J48graft pruned tree
-----

Categoria Administrativa = PÚBLICA: UNIVERSIDADES (937.0/54.0)
Categoria Administrativa = PRIVADA
| UF = MT
| | Conceito Enade = 5.0: FACULDADES (0.0)
| | Conceito Enade = 4.0
| | | Número de Participantes Concluintes <= 113.5
| | | | Número de Concluintes Inscritos <= 116.5
| | | | | Média FG padronizada <= 61.91025
| | | | | Média FG bruta <= 60.04555
| | | | | Nota Contínua <= 3.77255
| | | | | Média CE bruta <= 59.6932: CENTROS UNIVERSITÁRIOS (3.0/1.0)
| | | | | Média CE bruta > 59.6932: UNIVERSIDADES (0.0|43.0/6.0)
| | | | | Nota Contínua > 3.77255: UNIVERSIDADES (0.0|12.0/1.0)
| | | | | Média FG bruta > 60.04555: UNIVERSIDADES (0.0|31.0/3.0)
| | | | | Média FG padronizada > 61.91025: UNIVERSIDADES (0.0|7.0)
| | | | | Número de Concluintes Inscritos > 116.5: UNIVERSIDADES (0.0|8.05)
| | | | | Número de Participantes Concluintes > 113.5: UNIVERSIDADES (0.0|8.05)
| | Conceito Enade = SC
| | | Número de Concluintes Inscritos <= 3
| | | Média CE padronizada <= 24.1468
| | | | Média CE padronizada <= 21.34055: FACULDADES (0.0|14.0)
| | | | Média CE padronizada > 21.34055
| | | | | Média FG padronizada <= 40.77805
| | | | | Média FG padronizada <= 31.2195: FACULDADES (0.0|11.0/2.0)
| | | | | Média FG padronizada > 31.2195: UNIVERSIDADES (2.0/1.0)
| | | | | Média FG padronizada > 40.77805: FACULDADES (0.0|13.0/1.0)
| | | | Média CE padronizada > 24.1468: FACULDADES (24.0/4.0)

```

Figura 11. Trecho do Modelo gerado no J48GRAFT aplicado ao Arquivo ENADE 2012. Arff.

Com Instâncias classificadas corretamente em 90%, na árvore gerada pelo algoritmo, os dados apresentam padrões que nos dizem que entre as Universidades, Faculdades e Centros Universitários, nos estados de Mato Grosso, Distrito Federal, na região Centro Oeste do país, junto ao estado de Sergipe, no Nordeste, Amazonas e Piauí, na região Norte, instituições públicas, em geral, obtiveram resultados conclusivos no exame em 2012, com conceitos que variam, entre Excelente, Ótimo e Regular, sendo a última, a mais comum dentre o grupo. Enquanto que as instituições privadas nestes estados conseguiram conceitos Ruim e Péssimo, em maior ocorrência, assim como houve casos de não conclusão do exame, por estas mesmas instituições.

Nos estados de Minas Gerais e São Paulo, na região Sudeste, os desempenhos dentre as unidades classificadas como Públicas, tais como Universidades e Centros Universitários, obtiveram conceitos Ótimo e Excelente, em seu conceito ENADE. Já as instituições privadas pertencentes a estes estados, os conceitos variam entre Regular, Ruim e Péssimo. Sendo que o conceito Regular é a nota com maior ocorrência dentre estes grupo específico de unidades avaliadas.

Nos estados do Paraná, Rio Grande do Sul e Santa Catarina, na região Sul do país, junto aos estados de Rio de Janeiro, no Sudeste e Pernambuco, Bahia, Ceará, Rio Grande

do Norte, no Nordeste, e também, Goiás, no Centro Oeste, tiveram suas instituições públicas, avaliadas pelo exame, com resultados finais iguais a Ótimo, Excelente e Péssimo. Sendo que dentre este grupo de unidades, o conceito Ótimo, possui a maior ocorrência. Pode-se dizer que nas Faculdades, ou unidades privadas, obtiveram conceito Regular, sendo esta com maior frequência, junto a conceito Ruim, além de casos de não conclusão do exame.

Nos estados do Espírito Santo na região Sudeste, junto aos estados da Paraíba e Alagoas, na região Nordeste, Mato Grosso do Sul, na região Centro Oeste, e ainda, estados da região Norte do país, tais como Rondônia, Tocantins, Maranhão, Roraima, Amapá e Acre, possuem semelhanças dentre seus padrões, mostrando que as Universidades e Centros Universitários, obtiveram conceitos iguais a Ruim e Péssimo, notas que não demonstram desempenhos satisfatórios, em comparação as demais instituições deste mesmo grupo, nos demais estados descritos acima. Porém, pode-se observar que as Unidades Privadas, conseguiram notas melhores, que demonstram índices classificados Excelente, Ótimo e Regular, nestes estados.

#### 4.7. APLICAÇÕES DAS REGRAS DE ASSOCIAÇÃO COM ALGORITMO APRIORI

O algoritmo APRIORI, foi utilizado para gerar os principais padrões entre os dados que demonstre o desempenho das Instituições de Ensino Superior no exame do ENADE, relacionando-os com o censo do ENADE, e também, com a Nota Contínua. Neste sentido, as tabelas 14 e 15, destacam os principais padrões encontrados pelo algoritmo nos arquivos, CENSO\_ENADE 2013. Arff e CENSO\_ENADE 2013 - 2. Arff, representando ocorrências mais relevantes na avaliação realizada no ano de 2013. Esta aplicação serviu para traçar as informações do desempenho geral de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelas mesmas.

Tabela 14 – Regras de associação utilizando os atributos “Área de Enquadramento, Categ. Administrativa, Org. Acadêmica, Curso, Conceito ENADE, Nota Contínua Conceito ENADE”.

Suporte mínimo	0.15
Confiança Mínima de Padronização	1.0
Nº de ciclos	17
<b>Regras Geradas / CENSO_ENADE 2013. Arff</b>	<b>Confiança</b>
1. Nota Contínua Conceito ENADE=Bom 1373 ==> Conceito ENADE=regular 1373.	1.0
2. Conceito ENADE=regular 1373 ==> Nota Contínua Conceito ENADE=Bom 1373.	1.0
3. Nota Contínua Conceito ENADE=Regular 881 ==> Conceito ENADE=Ruim 881	1.0
4. Conceito ENADE=Ruim 881 ==> Nota Contínua Conceito ENADE=Regular 881.	1.0
5. Nota Contínua Conceito ENADE=Ótimo 808 ==> Conceito ENADE=Ótimo 808.	1.0
6. Conceito ENADE=Ótimo 808 ==> Nota Contínua Conceito ENADE=Ótimo 808.	1.0
7. Categ. Administrativa=Privada Sem Fins Lucrativos Nota Contínua Conceito ENADE=Bom 696 ==> Conceito ENADE=regular 696.	1.0
8. Categ. Administrativa=Privada Sem Fins Lucrativos Conceito ENADE=regular 696 ==> Nota Contínua Conceito ENADE=Bom 696.	1.0
9. Org. Acadêmica=Universidade Conceito ENADE=regular 626 ==> Nota Contínua Conceito ENADE=Bom 626.	1.0
10. Org. Acadêmica=Universidade Conceito ENADE=regular 626 ==> Nota Contínua Conceito ENADE=Bom 626.	1.0

As principais regras sendo elas, 8, 9 e 10 na Tabela 14, segundo o padrão de dados, mostram que as Faculdades, classificadas como privadas Sem Fins Lucrativos obtiveram bom desempenho na Nota Contínua do ENADE no ano de 2013, porém, seu conceito ENADE conclusivo, foi Regular. Enquanto que as Universidades Públicas, destacadas nas regras 9 e 10 na Tabela 14, obteve desempenho semelhante, sendo que suas notas contínuas no censo foram Boas, e a nota final do exame, como Regular. Junto a isso, pode-se conferir que houve Universidades Públicas, que obteve desempenho semelhante, sendo que suas notas contínuas no censo foram Regulares, e o conceito ENADE também classificado como Regular.

Nas regras 7, 8, 9 e 10 na Tabela 14 dizem também, que a maior parte das instituições, obteve conceito ENADE considerado como Regular. Sendo que significativa parcela das instituições, alcançaram resultados Ruins, no censo geral. Por outro lado, outra

parcela relativamente menor que a anterior, alcançou resultado, classificado como Ótimo. Pode-se concluir que o conceito Regular, segundo regras geradas obteve a maior ocorrência dentre as instituições avaliadas no exame do ano de 2013. Com tais padrões de dados, é possível observar com detalhes a Educação Superior no Brasil, que segundo os dados, não tem desempenhado trabalhos que levem aos melhores índices de desempenho e qualificação em avaliações institucionais, refletindo assim a qualidade do ensino ministrado pelas mesmas.

Tabela 15 – Regras de associação utilizando os atributos “*Categ. Administrativa, Org. Acadêmica, UF, Nota Padronizada – CE, Nota padronizada – FG, Nota Contínua Conceito ENADE*”.

Suporte mínimo	0.35
Confiança Mínima de Padronização	1.0
Nº de ciclos	13
<b>Regras Geradas / CENSO_ENADE 2013 – 2. Arff</b>	<b>Confiança</b>
1. Conceito ENADE=regular 1373 ==> Nota Padronizada - CE=Bom 1373	1.0
2. Nota Padronizada - CE=Bom 1373 ==> Conceito ENADE=regular 1373	1.0
3. Nota Contínua Conceito ENADE=Bom 1373 ==> Nota Padronizada - CE=Bom 1373	1.0
4. Nota Padronizada - CE=Bom 1373 ==> Nota Contínua Conceito ENADE=Bom 1373	1.0
5. Nota padronizada - FG=Bom 1373 ==> Nota Padronizada - CE=Bom 1373	1.0
6. Nota Padronizada - CE=Bom 1373 ==> Nota padronizada - FG=Bom 1373	1.0
7. Nota Contínua Conceito ENADE=Bom 1373 ==> Conceito ENADE=regular 1373	1.0
8. Conceito ENADE=regular 1373 ==> Nota Contínua Conceito ENADE=Bom 1373	1.0
9. Nota padronizada - FG=Bom 1373 ==> Conceito ENADE=regular 1373	1.0
10. Conceito ENADE=regular 1373 ==> Nota padronizada - FG=Bom 1373	1.0

As principais regras sendo elas 3, 4, 5, 6 e 9 na Tabela 15, diz que o conceito ENADE Regular foi alcançado por uma parcela considerável das instituições participantes do exame no ano de 2013, e por isso, a nota de Componente Específico, é considerada Boa. Com isso entende-se que, a formação oferecida em certos cursos, reflete diretamente no desempenho dos alunos no exame. E também, nas mesmas parcelas de instituições, as notas contínuas foram satisfatórias.

Ainda nas regras 3, 4, 5, 6 e 9 na Tabela 15, os dados relatam que uma parte das instituições alcançou no quesito Formação Geral, conceito Bom, junto à nota de Componente Específico. Porém, nesta mesma relação, há instituições que apesar de ter alcançado bom resultado na Nota Contínua, seu conceito final no exame, sofre uma queda, para conceito Regular. Nesta mesma parcela de instituições, a Formação Geral, alcançou Boa classificação, enquanto que o conceito final do exame foi Regular.

#### 4.8. APLICAÇÕES DA ÁRVORE DE DECISÃO COM ALGORITMO J48GRAFT

Em outra análise dos resultados, foi criada a Árvore de Decisão utilizando o algoritmo J48GRAFT. A seguir, são demonstrados os principais resultados provenientes, desta técnica de Mineração de Dados. Esta aplicação serviu para traçar as informações do desempenho geral de cada estado, categoria administrativa, e organização acadêmica, de Universidades e Faculdades avaliadas, expondo as notas gerais alcançadas pelos cursos, pertencentes às mesmas, através de um mapeamento.

```
J48graft pruned tree
-----
Categ. Administrativa = Pública Federal: Universidade (540.0/59.0)
Categ. Administrativa = Pública Estadual: Universidade (235.0/22.0)
Categ. Administrativa = Privada sem fins lucrativos: Universidade (1567.0/836.0)
Categ. Administrativa = Pública Municipal
| UF = MT: Faculdade (1.0)
| UF = DF
| | Conceito Enade = 4.0: Universidade (0.0|25.0/4.0)
| | Conceito Enade != 4.0: Faculdade (0.0)
| UF = PI
| | Conceito Enade = 5.0: Universidade (0.0|6.0)
| | Conceito Enade != 5.0
| | | Conceito Enade = 4.0: Universidade (0.0|27.0/7.0)
| | | Conceito Enade != 4.0: Faculdade (0.0)
| UF = MG: Faculdade (3.0)
| UF = PR
| | Conceito Enade = 5.0: Universidade (0.0|22.0/4.0)
| | Conceito Enade != 5.0: Faculdade (3.0/1.0)
| UF = RS
| | Conluintes Inscritos <= 131.5
| | | Conceito Enade = 5.0: Universidade (0.0|22.0/3.0)
| | | Conceito Enade != 5.0
| | | | Conceito Enade = 4.0: Universidade (0.0|89.0/19.0)
| | | | Conceito Enade != 4.0
| | | | | Conluintes Inscritos <= 3.5: Universidade (0.0|15.0/4.0)
| | | | | Conluintes Inscritos > 3.5
| | | | | | Conluintes Participantes <= 3.5: Universidade (0.0|17.0/5.0)
| | | | | | Conluintes Participantes > 3.5
| | | | | | Conceito Enade = 3.0: Universidade (0.0|158.0/50.0)
```

Figura 12. Trecho do Modelo gerado no J48GRAFT aplicado ao Arquivo ENADE 2013. Arff.

Com Instâncias classificadas corretamente em, aproximadamente 70 %, segundo o modelo gerado pelo Algoritmo J48GRAFT, pode-se concluir que: Referindo-se às Universidades Públicas Federais, o estado de Minas Gerais obteve o maior índice de bons resultados finais na avaliação dos cursos, sendo classificados como Ótimos e Excelentes conceitos. Em compensação, os cursos pertencentes nesta categoria administrativa (Universidade Pública), nos estados do Paraná, Minas Gerais, Rio de Janeiro, Mato Grosso, Rio Grande do Sul, Goiás, São Paulo, Pernambuco e Roraima, alcançaram resultado Regular no exame. O curso de educação física, especificamente, obteve conceito Regular, em toda a Organização acadêmica, nos estados de Goiás, Paraná e Minas Gerais. Parte dos cursos, de algumas instituições, pertencentes aos estados do Rio de Janeiro, Paraná e Mato Grosso do Sul, obtiveram conceito Péssimo. Pode-se observar que, os cursos de algumas instituições obtiveram conceito Ruim, nos estados do Pernambuco, Rio de Janeiro, Amazonas, Paraná, Bahia, Rio Grande do Sul, Goiás, Piauí, Acre e Minas Gerais.

Percebeu-se que a região Sudeste, com todas as instituições avaliadas no ano de 2013, obtiveram os melhores índices de desempenho no exame, obtendo conceitos Ótimo e Excelente, em especial no estado de Minas Gerais. Porém, houve parte de instituições que obtiveram conceito Regular no censo final. Enquanto a isso, as regiões Norte e Nordeste obtiveram os piores resultados, alcançando conceito Ruim e Péssimo.

Referindo-se as Universidades Públicas Estaduais, notou-se que os cursos avaliados no ano de 2013, obtiveram desempenhos no exame, de forma específica. Podem-se evidenciar ocorrências comuns nos dados observando alguns cursos avaliados na área de saúde, tais como:

No curso de Medicina Veterinária: conceitos Ótimo e Excelente nos estados de São Paulo e Paraná. Porém, no Paraná houve notas também classificadas como Regular e Péssimo, assim como instituições que não concluíram no exame.

No curso de Odontologia, nos estados da Bahia e Paraná, tiveram conceitos Excelente e Ótimo. Enquanto que no estado de São Paulo, foi alcançado conceito Regular. No curso de Medicina, apenas o estado do Paraná, alcançou conceito Excelente no exame.

Tanto instituições públicas, como privadas, no estado de São Paulo, tiveram conceito Ótimo, e também, casos de desempenho de conceitos Péssimos, e de não conclusão do exame, por algumas instituições.



No curso de Enfermagem, apenas no estado do Paraná, obteve conceito Excelente no curso em questão. Enquanto que instituições tanto públicas e privadas tiveram conceito Ótimo nos estados do Paraná e São Paulo. Enquanto que estados da região Norte e Nordeste, obtiveram conceitos Regulares e Ruins, no exame.

As instituições classificadas como Privadas Sem Fins Lucrativos, assim como as Públicas Municipais, tiveram mais participação no exame, nos estados de Santa Catarina e São Paulo. As instituições classificadas como Privadas com Fins Lucrativos, assim como Universidades e Faculdades do estado de São Paulo, tiveram casos específicos nos resultados dos cursos avaliados, tais como:

O curso de farmácia, nos estados de Paraná, e Goiás, obtiveram desempenhos Ruins no exame, enquanto que o estado de Minas Gerais, apesar de obter casos de não conclusão do exame, houve bom índice de desempenho, classificados como Ótimo e Excelente. O curso de serviço social, no estado do Paraná, aparece com índices de desempenho classificados como satisfatórios, porém, com casos de algumas instituições que obtiveram conceito Ruim, ao final do exame. Enquanto que o estado de Minas Gerais se destaca mais de uma vez, por seu desempenho, considerado Ótimo.

O curso de tecnologia em radiologia, no estado de Minas Gerais, alcançou resultados Ótimos e Excelentes, havendo casos também, de conceito Ruim, e casos de não conclusão. No Distrito Federal, o alcance de desempenhos foi classificado como Ruins e Regulares.

O curso de zootecnia, no estado do Maranhão ganha destaque, pois conseguiu os melhores índices de aprovação no exame, alcançando resultados Excelente e Ótimo, porém, em contrapartida, houve casos classificados como Ruins. E no estado de Roraima, o conceito Ruim, passa a ser o grupo de maior ocorrência neste curso.

Pode-se concluir que os melhores desempenhos, foram alcançados pelos estados de Minas Gerais e Paraná, pertencentes a regiões do Sudeste e Sul. E também, que os cursos dos estados do Norte e Nordeste, em áreas especializadas no Campo, ganha maior representatividade, nem sempre satisfatória. As diferentes instituições Públicas e Privadas, com Categoria Administrativa, e Organização Acadêmica, distintas, tiveram equilíbrio, em relação do aos cursos avaliados, sendo que alguns obtinham resultado Excelente, enquanto que outros tiveram resultados classificados abaixo da linha de aprovação, havendo casos ainda, de não conclusão do exame, conforme o modelo gerado.

## 4.9. ESTUDO COMPARADO DO DESEMPENHO GERAL NO ENADE

### 4.9.1. Instituições Privadas

Tabela 16 – Dados gerais de desempenho (Privadas).

Ano	Desempenho Geral (Nota ENADE)
2010	Regular
2011	Regular
2012	Regular
2013	Regular

Na Tabela 16, são apresentados os dados gerais do desempenho das IEs Privadas. É possível observar que as Instituições classificadas como privadas (Faculdades), desde o ano de 2010, tem mantido o mesmo resultado conclusivo geral de desempenho no exame. Isso mostra uma estagnação do desenvolvimento educacional, e da formação técnica dos cursos avaliados destas instituições, sendo estes, os principais atributos de avaliação do exame. Os dados mostram que durante esses quatro anos, a qualidade educacional em Faculdades privadas em todo país, não sofreu mudanças que apoiassem qualitativamente a formação oferecida para os alunos, que são refletidas no exame, já que estes demonstram desempenho Regular na conclusão do curso. Na Figura 13 é apresentado o gráfico onde se pode constatar que, as instituições privadas que realizaram o exame do ENADE, dentre os anos 2010 a 2013, tiveram o mesmo desempenho geral na conclusão do exame.

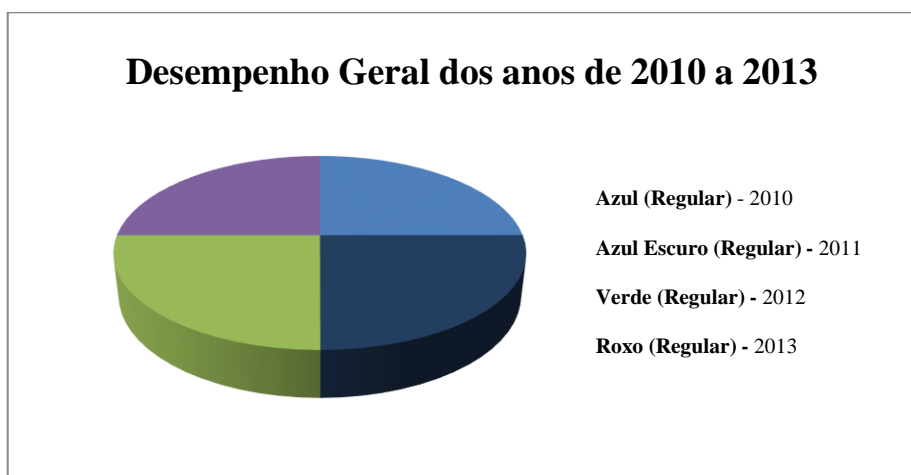


Figura 13. Modelo Gráfico do desempenho geral de Instituições Privadas.

O gráfico acima, mostra de forma simples, que o desempenho das instituições privadas, desde ao ano de 2010, ao ano de 2013, vem mantendo-se estático, alcançando somente classificações Regulares.

#### 4.9.2. Instituições Públicas

Tabela 17 – Dados gerais de desempenho (Públicas).

<b>Ano</b>	<b>Desempenho Geral (Nota ENADE)</b>
2010	Regular
2011	Ótimo
2012	Ruim
2013	Regular

Na Tabela 17, são apresentados os dados gerais de desempenho das instituições públicas. Pode-se observar que as Instituições classificadas como Públicas, em todo o país, desde o ano de 2010, tem mantido certa oscilação de resultados conclusivos de desempenho no exame. Isso mostra uma variação considerável do desenvolvimento educacional, e da formação técnica dos cursos avaliados destas instituições, sendo estes os principais atributos de avaliação do exame, dito anteriormente. Os dados mostram que durante esses quatro anos, a qualidade educacional em Universidades, assim como as instituições particulares, não sofreram mudanças que apoiassem qualitativamente a formação oferecida para os alunos, já que estes ao final do exame demonstram desempenho Regular na conclusão do curso, de forma predominante. Apenas em exceção do ano de 2011, que se mostrou com melhor resultado final, e o pior em 2012, como conceito Ruim. Com o apoio do gráfico abaixo, pode-se então constatar que, as instituições públicas que realizaram o exame do ENADE, dentre os anos de 2010 a 2013, tiveram diferentes desempenhos gerais na conclusão do exame, sendo Bons e Ruins.

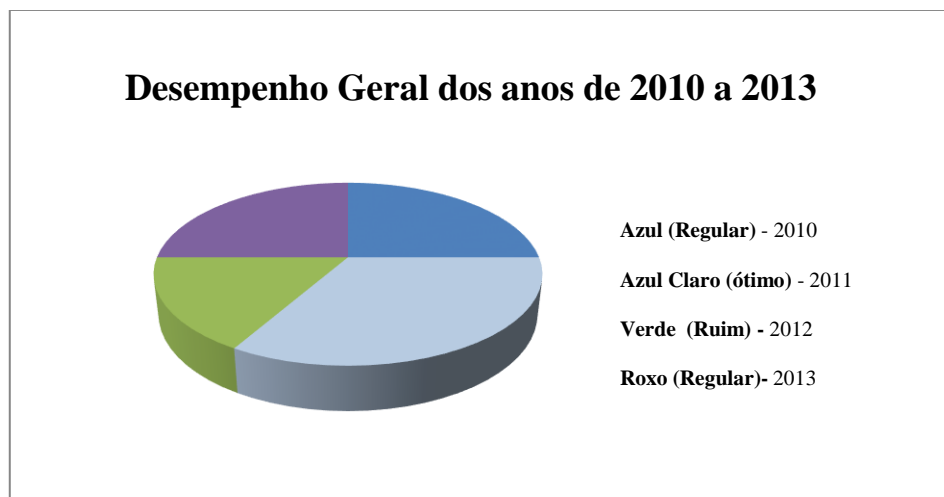


Figura 14. Modelo Gráfico do desempenho Geral de Instituições Públicas.

O gráfico acima, na Figura 14, mostra de forma simples, que o desempenho das instituições de ensino público, desde ao ano de 2010, ao ano de 2013, vem sofrendo variações, em classificações que vão desde Ruins a conceitos Ótimos.

## 5. CONCLUSÃO

Com o crescente desenvolvimento do armazenamento de dados em bases eletrônicas, faz-se necessária a obtenção de métodos que possam ajudar a interpretar o que tais dados podem trazer de útil para empresas, e as demais entidades detentoras de grandes massas de dados. Com o sistema educacional não é diferente. Faz-se necessário que técnicas, como a Mineração de Dados, possam está à disposição para que sejam aplicadas a essas grandes bases de dados, para entendermos a real situação do sistema educacional do Brasil. Nesta pesquisa, pôde-se observar qual a situação de qualidade do Ensino Superior ministrado, em cada região do país, que são mostrados pelo exame ENADE a cada ano de realização. A grande massa de Dados Educacionais tem um grande poder de levar a tomadas de decisão cada vez mais eficazes para a uma verdadeira renovação de regras aplicadas ao ensino ministrado em cursos pertencentes nas instituições de Ensino Superior. Assim, a qualidade é um recurso que tange o aperfeiçoamento não somente dos serviços e produtos (no caso os profissionais de educação e cursos), mas também da oferta, procura e adequação da própria satisfação do consumidor (alunos).

Segundo Silva (2012), o processo de Mineração de Dados é a extração de conhecimentos inesperados e não induzidos, ou seja, ela é responsável por aperfeiçoar o processo de geração de novos conhecimentos que levem a melhores tomadas de decisões, que levem a resolução de ocorrências adversas ao campo em que os dados pertencem, sendo essas, expostas por sua padronização. E a prática de análise de dados por Mineração de Dados, tem crescido cada vez mais, devido ao fato de que o Weka se consolidou como a ferramenta de Mineração de Dados mais utilizada em ambiente acadêmico para fins didáticos, e pesquisas científica.

O acesso a novos conhecimentos é capaz de modificar por completo a relação da universidade em um meio social. Tendo a real situação ao alcance de todos, pode-se tirar proveito para encontrar as soluções para tantas deficiências existentes em meio ao Sistema Educacional Brasileiro, como o fato de tantas instituições, que de acordo com o estudo, apenas alcançam resultados Regulares, no longo dos últimos anos, assim como, tantas instituições das regiões Norte e Nordeste, atingiram resultados mínimos em seus resultados finais, considerados Ruins e Péssimos, demonstrando grande impacto deficiente da qualidade da formação oferecida nestas instituições, apresentado neste trabalho. De forma mais específica, vê-se que o desempenho das instituições privadas, desde ao ano de 2010

ao ano de 2013, permaneceu estagnado, em classificações Regulares, enquanto que as Universidades Públicas tem seu desempenho, desde ao ano de 2010 a 2013, sofre variações, em classificações que vão desde Ruins a conceitos Ótimos.

Para futuros trabalhos, há a pretensão de buscar bases de dados relacionadas à educação básica (ensino fundamental e médio), com o intuito de aplicar a mesma metodologia de pesquisa para que assim, da mesma forma, sejam estabelecidos padrões para a geração de conhecimento útil, com a Mineração de Dados.

Segundo Vermelho (1998), a formação do atual sistema de ensino básico nas escolas, com metodologias que já são praticamente ineficazes em sala de aula, fruto da educação tradicional, o processo ensino-aprendizagem, ganha possibilidades de apresentar novas formas de aperfeiçoamento, e ao mesmo tempo, torne o ambiente escolar, algo interessante e estimulante, apenas com o uso de ferramentas que proporcionem a inserção de tecnologias para o auxílio do processo de aprendizagem nas escolas. E sob esta perspectiva, que se dará a construção desta nova base de dados, para a submissão destes dados a serem coletados, para a mineração. E por meio dos novos conhecimentos, tal objetivo é realizar a criação de métodos que possam ser úteis para a formulação, em longo prazo, de processos de treinamento e a capacitação de vários profissionais de educação da região, para que assim o modelo educacional presente, possa torna-se mais eficaz e desenvolvido, tendo a tecnologia, uma ferramenta de apoio e suporte a decisão, a partir das decisões e planos traçados, após a Mineração de Dados.

Voltando a pesquisa já desenvolvida, a partir do conhecimento descrito neste trabalho, a realização da tomada de decisão pelos os órgãos responsáveis pelo exame do ENADE, como o INEP e o MEC, a metodologia baseada na Mineração de Dados levam esses Dados Educacionais a terem a possibilidade de melhor processamento e análise, para identificar e solucionar falhas existentes dentre o Sistema de Ensino Superior no Brasil. Tendo o real conhecimento da qualidade educacional refletida pelos resultados do exame, tais órgãos agora, podem tirar proveito para construir novos métodos que levem as instituições de Ensino Superior, demonstrarem melhor qualidade, e principalmente, melhor desempenho em avaliações institucionais, em todo território nacional.

## REFERENCIAS

AMO, Sandra de. **Técnicas de Mineração de Dados**. Uberlândia: Universidade Federal de Uberlândia Faculdade de Computação, 2003. 43 p.

ANDRADE, Cibele Yahn de. **Acesso ao Ensino Superior no Brasil: equidade e desigualdade social**. Ensino Superior Unicamp, São Paulo, p.1-12, jul. 2012. Disponível em: <[HTTPS://www.revistaensinosuperior.gr.unicamp.br/artigos/acesso-ao-ensino-superior-no-brasil-equidade-e-desigualdade-social](https://www.revistaensinosuperior.gr.unicamp.br/artigos/acesso-ao-ensino-superior-no-brasil-equidade-e-desigualdade-social)>. Acesso em: 26 mar. 2015.

BAKER, Ryan Shaun Joazeiro de; ISOTANI, Seiji; CARVALHO, Adriana Maria Joazeiro Baker de. **Mineração de Dados Educacionais: Oportunidades para o Brasil**. 22. Ed. Worcester: Department Of Social Sciences And Policy Studies, 2011. 12 p. Disponível em: <<http://users.wpi.edu/~rsbaker/BD-RBIE-pt-v22.pdf>>. Acesso em: 18 mai. 2015.

BOSCHI, Raquel S.; OLIVEIRA, Stanley R. de M.; ASSAD, Eduardo D.. **Técnicas De Mineração de Dados Para Análise Da Precipitação Pluvial Decenal No Rio Grande Do Sul**. Scielo, Campinas, v. 31, n. 6, p.1-13, dez. 2011. Disponível em: <<http://www.scielo.br/pdf/eagri/v31n6/v31n6a16.pdf>>. Acesso em: 22 jun. 2015.

BRAGA, Ronald. **Qualidade e eficiência do modelo de Ensino Superior brasileiro: uma reflexão crítica**. São Paulo: Núcleo de Pesquisas Sobre Ensino Superior, 2010. 17 p. Disponível em: <<http://nupps.usp.br/downloads/docs/dt8910.pdf>>. Acesso em: 27 mar. 2015.

CAMPOS, Maria. **A pesquisa em Educação no Brasil**. São Paulo: Fundação Carlos Chagas – PUC/SP, 1994. 13 p.

CAMILO, Cássio Oliveira; SILVA, João Carlos Da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. 9. Ed. Goiânia: Instituto de Informática/UFG, 2009. 29 p. Disponível em: <[http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-09.pdf](http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf)>. Acesso em: 03 mai. 2015.

COSTA, Claudio Napolis; COUTINHO, Jonatas Vieira; MAGALHÃES, Lúcia Helena de. **DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS. Fundação Educacional São José: Revista Eletrônica**, Santos Dumont, v. 2, n. 2, p.1-10, 2005. Semanal. Disponível em: <<http://fsd.edu.br/revistaeletronica/arquivos/2Edicao/artigo9.pdf>>. Acesso em: 28 abr. 2016.

CUNHA, Luiz Antônio. **Desenvolvimento Desigual E Combinado No Ensino Superior – Estado E Mercado**. Scielo, Campinas, v. 25, n. 1, p.1-23, out. 2004. Disponível em: <<http://www.scielo.br/pdf/ES/v25n88/a08v2588.pdf>>. Acesso em: 22 jun. 2015.

DIAS, Maria Madalena. **Um Modelo De Formalização Do Processo De Desenvolvimento De Sistemas De Descoberta De Conhecimento Em Banco De Dados**. 2012. 212 f. Tese (Doutorado) - Curso de Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2012. Cap. 02. Disponível em: <[http://www.din.uem.br/~mmdias/documentos/tese\\_Madalena.pdf](http://www.din.uem.br/~mmdias/documentos/tese_Madalena.pdf)>. Acesso em: 05 abr. 2015.

EVAN, Robert. **Do Provão ao ENADE: uma análise comparativa dos exames nacionais utilizados no Ensino Superior Brasileiro**. Rio de Janeiro: INEP, 2006. 19 p.

FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **Advances in Knowledge Discovery and Data Mining**. Menlo Park, EUA: AAAI Press, 1996. 611 p.

FONSECA, Denise Grosso da; FORSTER, Mari Margarete dos Santos. **Implicações Do Exame Nacional De Desempenho Dos Estudantes (ENADE) Nos Processos Avaliativos Internos De Um Curso De Educação Física**. Anpae, São Leopoldo, v. 1, n. 1, p.1-18, nov. 2007. Mensal. Disponível em: <[http://www.anpae.org.br/congressos\\_antigos/simposio2007/111.pdf](http://www.anpae.org.br/congressos_antigos/simposio2007/111.pdf)>. Acesso em: 19 jun. 2015.

GALVÃO, Noemi Dreyer; MARIN, Heimar de Fátima. **Técnica de Mineração de Dados: uma revisão da literatura**. Scielo, São Paulo, v. 1, n. 1, p.1-5, abr. 2009. Disponível em: <<http://www.scielo.br/pdf/ape/v22n5/14.pdf>>. Acesso em: 22 jun. 2015.

HILU, Luciane. **Produção científica no Brasil - Um comparativo entre as Universidades públicas e privadas**. Paraná: Pontifícia Universidade Católica do Paraná, 2011. 9 p.

LIMA, Licínio C.; AZEVEDO, Mário Luiz Neves de; CATAN, Afrânio Mendes. **O Processo De Bolonha, A Avaliação Da Educação Superior E Algumas Considerações Sobre A Universidade Nova**. Scielo, Campinas, v. 13, p.1-31, mar. 2008. Disponível em: <<http://www.scielo.br/pdf/aval/v13n1/a02v13n1>>. Acesso em: 23 jun. 2015.



LONGO, Waldimir Pirró e. Alguns impactos sociais do desenvolvimento científico e tecnológico. **Datagramazero: Revista de Ciência da Informação**, São Paulo, v. 8, n. 1, p.01-20, fev. 2007. Mensal. Disponível em: <[http://www.dgz.org.br/fev07/Art\\_03.htm](http://www.dgz.org.br/fev07/Art_03.htm)>. Acesso em: 08 abr. 2015.

MARTINS, António Cardoso; MARQUES, João Miguel; COSTA, Paulo Dias. **Estudo Comparativo de Três Algoritmos de Learning na Classificação de Dados Eletrocardiográficos**. 2009. 16 f. Dissertação (Mestrado) - Curso de Medicina, Faculdade de Medicina da Universidade do Porto, Porto, 2009. Disponível em: <[https://www.dcc.fc.up.pt/~ines/aulas/0910/MIM/trabs\\_ano\\_anterior/noname-1.pdf](https://www.dcc.fc.up.pt/~ines/aulas/0910/MIM/trabs_ano_anterior/noname-1.pdf)>. Acesso em: 24 mar. 2016.

MENDES, Afrânio. A EDUCAÇÃO SUPERIOR NO BRASIL E AS TENDÊNCIAS DAS POLÍTICAS DE AMPLIAÇÃO DO ACESSO. **Atos de Pesquisa em Educação**, Blumenau, v. 2, n. 3, p.1-16, 2007. Semanal. Disponível em: <<http://proxy.furb.br/ojs/index.php/atosdepesquisa/article/view/754/630>>. Acesso em: 15 out. 2015.

NAMEN, Anderson Amendoeira; BORGES, Sonia Xavier de Almeida; SADALA, Maria da Glória Schwab. **Indicadores de qualidade do ensino fundamental: O uso das tecnologias de Mineração de Dados e de visões multidimensionais para apoio à análise**.

NAVEGA, Sergio. **Princípios Essenciais da Mineração de Dados**. In: Infoimagem 2002, 07, 2002, São Paulo. Anal. São Paulo: Cenadem, 2002. P. 1 - 9. Disponível em: <<http://www.intelliwise.com/reports/i2002.pdf>>. Acesso em: 03 abr. 2015.

OLIVEIRA, Julio. **Saiba como funciona sistema de Ensino Superior no Brasil**. 2009. Disponível em: <<http://www.brasil.gov.br/educacao/2009/11/ensino-superior>>. Acesso em: 26 mar. 2015.

PEDROSA, Renato Hyuda de Luna. **Um Plano Diretor para o Ensino Superior Público do Estado de São Paulo**. Ensino Superior Unicamp, São Paulo, v. 02, n. 02, p.1-18, nov. 2010. Quadrimestral. Disponível em: <[http://www.gr.unicamp.br/ceav/revistaensinosuperior/ed02\\_novembro2010](http://www.gr.unicamp.br/ceav/revistaensinosuperior/ed02_novembro2010)>. Acesso em: 17 jun. 2015.

QUINTELLA, Rogério Hermida; SOARES JUNIOR, Jair Sampaio. **Sistemas De Apoio À Decisão E Descoberta de Conhecimento em Bases De Dados: Uma Aplicação Potencial em Políticas Públicas**. Scielo, Salvador, v. 10, n. 28, p.1-16, dez. 2003. Disponível em: <<http://www.scielo.br/pdf/osoc/v10n28/06.pdf>>. Acesso em: 23 jun. 2015.

ROTHEN, José Carlos; BARREYRO, Gladys Beatriz. **Expansão Da Educação Superior No Brasil E Avaliação Institucional: Um Estudo Do Sistema Nacional De Avaliação Da Educação Superior(SINAES)**. Revista Avaliação, Brasília, v. 1, n. 30, p.1-16, jul. 2010. Mensal. Disponível em: <<http://www.serie-estudos.ucdb.br/index.php/serie-estudos/article/view/144/193>>. Acesso em: 19 jun. 2015.

SILVA, Alcione Dias da ET al. **Mineração de Dados Aplicada À Base De Dados Tributária Municipal**. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 19, 2012, Cachoeiro de Itapemirim. Anais... . Bauru: Dataci, 2012. P. 1 - 14.

SILVA, Renata Mesquita da. **Avaliação Do Desempenho Docente, Segundo A Percepção Discente, Através De Mineração De Dados**. 2010. 61 f. Dissertação (Mestrado) - Curso de Mestrado Profissional em Pesquisa Operacional e Inteligência Computacional, Universidade Candido Mendes, Campos dos Goytacazes, 2010.

TROW, M. **Reflections on the transition from elite to mass to universal access: forms and phases of higher education in modern societies since WWII**. Berkeley: University of Califórnia, 2005.

VERMELHO, Cristina; BRITO, Glaucia da Silva; PURIFICAÇÃO, Ivonélia da. **Capacitação de professores para uso da tecnologia da informação no ambiente escolar da rede pública**. Brasília: Iv Congresso RIBIE, 1998. 06 p.