CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS
Recredenciado pela Portaria Ministerial nº 1.162, de 13/10/16, D.O.U nº 198, de 14/10/2016
ASSOCIAÇÃO EDUCACIONAL LUTERANA DO BRASIL
Alexandre Moraes Matos

MÓDULO PARA PLATAFORMA DE MINERAÇÃO E APRESENTAÇÃO DOS DADOS DO ENADE DA ÁREA DE PSICOLOGIA DOS ANOS DE 2009 A 2015

Palmas – TO
2019

Alexandre Moraes Matos

MÓDULO PARA PLATAFORMA DE MINERAÇÃO E APRESENTAÇÃO DOS DADOS DO ENADE DA ÁREA DE PSICOLOGIA DOS ANOS DE 2009 A 2015

Trabalho de Conclusão de Curso (TCC) II elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Me Heloise Acco Tives Leão.

Palmas – TO

2019

Alexandre Moraes Matos

MÓDULO PARA PLATAFORMA DE MINERAÇÃO E APRESENTAÇÃO DOS DADOS DO ENADE DA ÁREA DE PSICOLOGIA DOS ANOS DE 2009 A 2015

Trabalho de Conclusão de Curso (TCC) II elaborado e apresentado como requisito parcial para obtenção do título de bacharel em Sistemas de Informação pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. Me Heloise Acco Tives Leão

Aprovado em: _____/_____/_____

BANCA EXAMINADORA

_____

Prof. Me Heloise Acco Tive Leão

Orientador

Centro Universitário Luterano de Palmas – CEULP

_____

Prof. Me Fabiano Fagundes

Centro Universitário Luterano de Palmas - CEULP

_____

Prof. Dra. Irenides Teixeira

Centro Universitário Luterano de Palmas - CEULP

Palmas – TO

2019

AGRADECIMENTOS

Primeiramente a Deus, que me deu força, coragem e saúde para chegar ao fim desta jornada.

Agradeço à minha família que sempre me apoiou e fez de tudo para que fosse possível alcançar o objetivo da graduação.

Aos professores do curso de computação do CEULP/ULBRA que sempre fazem o que está ao alcance para ajudar seus alunos, desde os primeiros dias de aula.

À minha professora e orientadora Prof.ª. Me Heloise Acco Tive Leão que ajudou do início ao fim do projeto, dando total atenção e tirando dúvidas a qualquer momento, até mesmo respondendo aos vários e-mails sempre de forma rápida.

À Prof.ª. Dra. Irenides Teixeira, coordenadora do curso de Psicologia do CEULP/ULBRA, que se dispôs totalmente a ajudar na compreensão de informações sobre a área, tornando possível o progresso do trabalho.

A todos os que contribuíram, de alguma forma, para o alcance dos objetivos propostos neste trabalho.

RESUMO

MATOS, Alexandre Moraes. **Módulo para plataforma de mineração e apresentação dos dados do ENADE da área de Psicologia dos anos de 2009 a 2015**. 2019. 63 f. Trabalho de Conclusão de Curso (Graduação) – Curso de Sistemas de Informação, Centro Universitário Luterano de Palmas, Palmas/TO, 2019[1].

As grandes massas de dados que aumentam a cada dia vêm despertando interesse de pesquisadores em utilizar técnicas computacionais na tentativa de descoberta de novos conhecimentos. O conceito de mineração de dados surgiu e é frequentemente utilizado como base de apoio destas atividades de exploração. A diversidade de recursos, técnicas, algoritmos e modelos de processos existentes fazem com que a mineração de dados possa atender diferentes tipos de necessidades, independente do contexto na qual será aplicada. O presente trabalho teve por objetivo utilizar os recursos envolvidos na mineração de dados a fim de encontrar informações relevantes a partir dos microdados do ENADE, com foco nas respostas e questionários de informações pessoais dos estudantes da área de Psicologia dos anos de 2009, 2012 e 2015. Foram utilizadas tarefas de associação e agrupamento nas análises dos dados. Após a validação dos resultados encontrados, foi integrado um módulo na plataforma de apresentação de dados minerados do ENADE, o qual permite a visualização dos resultados obtidos através de gráficos. O projeto foi desenvolvido com passos do modelo de referência do CRISP-DM, seguindo adaptações em sua estrutura, para melhor atendimento dos resultados esperados. Com os resultados obtidos, é possível obter informações sobre o desempenho de estudantes da Psicologia divididos por área de conhecimento, e também informações sobre os perfis dos estudantes instituições de ensino com melhor avaliação no ENADE.

Palavras-chave: Mineração de Dados. CRISP-DM. ENADE. Algoritmos de Mineração de Dados. Psicologia.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

IES = Instituição de Ensino Superior

CPC = Conceito Preliminar do Curso

CRISP-DM – Cross Industry Standard Process for Data Mining

ENADE – Exame Nacional de Desempenho de Estudantes

Inep – Instituto Nacinal de Estudos e Pesquisas Educacionais

KDD – Knowledge Discovery in Databases

SUMÁRIO

1 INTRODUÇÃO

Desde o ano de 2004 ocorre no Brasil o Exame Nacional de Desempenho de Estudantes (ENADE) com objetivo de medir a qualidade dos cursos das instituições de ensino superior. O Inep (2015) afirma que o "ENADE avalia o rendimento dos concluintes dos cursos de graduação em relação aos conteúdos programáticos, habilidades e competências adquiridas em sua formação". Os dados dos exames realizados podem ser acessados de forma gratuita diretamente no portal do Inep.

Isotani e Bittencourt (2015) dizem que dados abertos são dados que têm sua utilização permitida de forma livre, podendo ser redistribuídos por quem quer que seja. Quem utiliza desses dados tem, no máximo, a obrigação de citar a fonte original dos mesmos. Os autores ainda explicam que a disponibilização de dados abertos é classificada com base no "Sistema de 5 Estrelas", onde, quanto maior o número de estrelas, maior é a visibilidade desses dados.

O Inep disponibiliza seus microdados gerados por avaliações, pesquisas e exames abertamente em seu próprio portal. Estão disponíveis dados de diferentes programas existentes no Brasil e suas informações servem de base para muitos estudos que ocorrem atualmente.

Da Silva, Boscarioli e Peres (2003) afirmam que, com o decorrer do tempo, a análise de informações tem se tornado um fator cada vez mais importante para empresas, servindo como forma de manterem-se no mercado, buscarem inovação, terem auxílio em tomadas de decisão e apoio no gerenciamento de negócios, resultando no alcance desejado de resultados. Com o objetivo de realizar análises de dados de forma automatizada, surgiu a mineração de dados, que fornece diversos recursos eficientes que podem ser utilizados em diferentes contextos.

Este trabalho apresentou a aplicação de técnicas de mineração de dados nos dados das respostas do ENADE, mais especificamente das provas do curso de Psicologia que ocorreram nos anos de 2009, 2012 e 2015, com objetivo de encontrar informações relevantes, possíveis padrões de respostas e identificação de perfis dos estudantes de acordo com seus desempenhos. Como base para aplicação das técnicas e orientação nas etapas do trabalho, foi utilizado o modelo CRISP-DM. Além disso, é apresentada a forma como foi o desenvolvimento de um módulo para plataforma de apresentação dos resultados obtidos com a aplicação das técnicas.

O trabalho desenvolveu-se sob a hipótese de que a mineração de dados das provas do ENADE permite obter informações sobre o desempenho dos estudantes da área de Psicologia e observar seus desempenhos em determinadas áreas do campo de estudo. Portanto, foram

definidos os objetivos de desenvolver um módulo para plataforma de visualização – plataforma já existente com dados minerados da área de computação - com dados minerados do ENADE da área de Psicologia dos anos de 2009 a 2015, aplicar técnicas e algoritmos de mineração de dados para tratamento e obtenção de resultados com os microdados do ENADE disponibilizados pelo Inep e visualizar o desempenho e os perfis dos estudantes dos estudantes com base nos dados obtidos da área de Psicologia.

2 REFERENCIAL TEÓRICO

Para compor a base teórica deste trabalho, foram abordados conceitos mineração de dados, suas técnicas e tarefas, bem como alguns dos algoritmos mais utilizados em suas aplicações. Também foi abordado o modelo de inferência CRISP-DM, metodologia voltada para processos de mineração de dados.

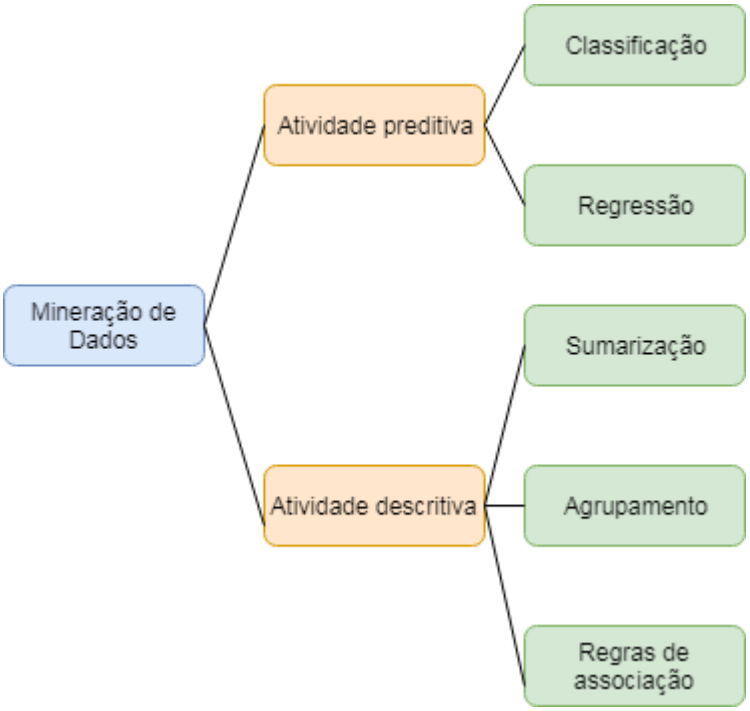
2.1 MINERAÇÃO DE DADOS

Mineração de dados pode ser entendida com a aplicação de técnicas automatizadas como forma de encontrar padrões e relações em grandes quantidades de dados que, a olho nu, não seriam possíveis de realizar (CARVALHO, 2001). Sua definição pode sofrer variações dependendo da área de atuação do autor, uma vez que suas técnicas podem ser aplicadas em diversas áreas.

Em finanças, a mineração de dados pode ser útil para estudo de tendências e análise de mercado, na área da saúde, pode-se usufruir de recursos para análise de perfis de pacientes, categorização de doenças ou estudos estatísticos. As técnicas de mineração podem ser aplicadas em qualquer conjunto de dados, não importando a área da qual os dados pertencem, porém é preciso saber quais tipos de resultados deseja-se obter a partir destes.

Amo (2004) ressalta que é importante diferir o que são técnicas e o que são tarefas de mineração de dados. Segundo esse autor, tarefa é o que se espera encontrar nos dados, quais tipos de informações e padrões podem ser alcançados em cima destes, já as técnicas são as identificações de métodos que permitem encontrar o que é almejado, como por exemplo técnicas de estatística e técnicas de aprendizado e máquina. A Figura 1 demonstra como são divididas as tarefas de mineração de dados.

Figura 1 - Divisão das tarefas de mineração de dados



Fonte: Adaptado de Camilo e Silva (2009)

Na figura acima é possível ver como estão divididas as principais áreas da mineração de dados. Na seção 2.1.1 estão detalhas tarefas de classificação e agrupamento, pois estas serão as tarefas utilizadas no desenvolvimento prático do trabalho. Primeiramente, realizando-se uma mineração não-supervisionada com a tarefas de agrupamento e, por seguinte, se os resultados encontrados permitirem, serão feitos estudos de forma supervisionada com a tarefa de classificação.

2.1.1 Classificação

Voznika e Viana (2007) afirmam que a tarefa de classificação consiste em prever as possíveis saídas de um determinado conjunto de dados de entrada. Com o objetivo de prever a saída, o algoritmo processa um conjunto de treinamento (*training set*) contendo os dados e sua respectiva saída, estes são comumente chamados de objetivo ou atributo de predição. Os autores afirmam que o algoritmo tenta descobrir relações entre os atributos, o que tornaria possível previr as saídas, o próximo passo seria fornecer ao algoritmo um conjunto de dados não observado antes, que é o conjunto de predição (*prediction set*), este conjunto contém os mesmos dados do conjunto de teste, porém, este não carrega a informação objetivo.

Tabela 1 - *Training set* de pacientes

Idade	Frequência cardíaca	Pressão sanguínea	Problema cardíaco
65	78	150/70	Sim
37	83	112/76	Não
71	67	108/65	Não

Fonte: Adaptado de Voznika e Viana (2007)

Na

Tabela 1 é possível ver um conjunto de características de pacientes e também a informação se os mesmos possuem problema cardíaco ou não. Esta tabela representa o *training set*. A Tabela 2 representa o *prediction set*, que contém os mesmos dados apresentados na

Tabela 1, mas sem a informação se o paciente possui problema cardíaco ou não.

Tabela 2 - *Prediction set* de pacientes

Idade	Frequência cardíaca	Pressão sanguínea	Problema cardíaco
43	98	147/89	?
65	58	106/63	?
84	77	150/65	?

Fonte: Adaptado de Voznika e Viana (2007)

Diante dos dados apresentados acima, o algoritmo de classificação de dados usaria as informações apresentadas na

Tabela 1 como um modelo, de forma a aprender com os dados ali dispostos e, a partir destes, determinar as chances de os pacientes listados na Tabela 2 terem, ou não, problemas cardíacos.

> "As técnicas de classificação podem ser supervisionadas e não-supervisionadas. São usadas para prever valores de variáveis do tipo categóricas. Pode-se, por exemplo, criar um modelo que classifica os clientes de um banco como especiais ou de risco, um laboratório pode usar sua base histórica de voluntários e verificar em quais indivíduos uma nova droga pode ser melhor ministrada. Em ambos os cenários um modelo é criado para classificar a qual categoria um certo registro pertence: especial ou de risco, voluntários A, B ou C" (CAMILO; SILVA, 2009).

Um dos algoritmos mais antigos e simples da tarefa de classificação é o K-Nearest Neighbor que, de acordo com Saini, Singh e Khosla (2013), é um algoritmo que trabalha com um método baseado em aprendizagem, gerando novas informações a partir de um conjunto de dados de treinamento.

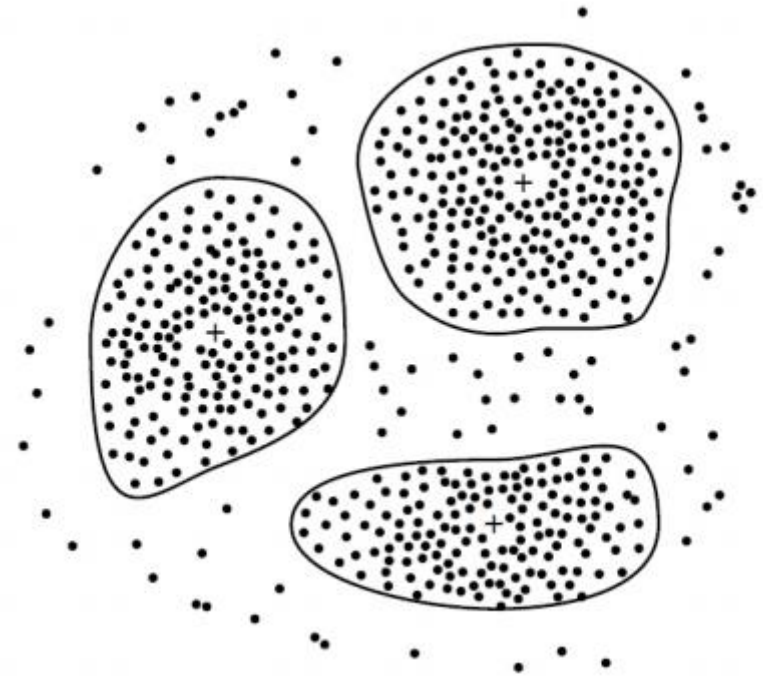
Os pesquisadores afirmam que o processo do algoritmo está dividido em duas etapas: na primeira, o algoritmo armazena os dados amostrais em forma de vetores e, na segunda fase, a de classificação, são selecionado os primeiros *K* (valor definido pelo usuário) elementos do vetor, que servem como base para a classificação, e os compara com os elementos da base de dados original, desta forma, o algoritmo define as distância de relação entre os elementos comparados.

2.1.2 Agrupamento

A tarefa de agrupamento, também conhecida como clusterização, tem como característica realizar a aproximação de dados similares. Camilo e Silva (2009) definem que um grupo (*cluster*) é um conjunto de registros diferentes, porém similares. Os pesquisadores afirmam que as possibilidades para o uso desta tarefa são inúmeras, sendo possível trabalhar com ela desde simples análises de dados até processamento de imagens.

Esta tarefa possui semelhanças com a tarefa de classificação, porém, Da Costa Côrtes, Porcaro e Lifschitz (2002) afirmam que a principal diferença entre as duas se dá no fato de que, no agrupamento, os conjuntos de registros ainda não estão categorizados, ou seja, o estudo será feito de forma não supervisionada. A Figura 2 apresenta centenas de registros que estão agrupados em três conjuntos de dados similares.

Figura 2 - Agrupamento de registros



Fonte: Camilo e Silva (2009)

A Tabela 3 abaixo simula a disposição de tipos de livros em prateleiras de uma biblioteca. Percebe-se que os gêneros dos livros estão nas prateleiras de forma misturada, dificultando para um leitor que deseja encontrar livros de um gênero específico ou semelhante.

Tabela 3 - Gêneros de livros desagrupados

Prateleira 1	Esportes
	Tecnologia
	Geografia
Prateleira 2	História
	Ficção Científica
	Religião

Com as aplicações de recursos da tarefa de agrupamento, é possível verificar as similaridades entre os gêneros e tipos de livros, e desta forma organizar a biblioteca para que a disposição fique mais bem alocada. A Tabela 4 apresenta as prateleiras após a utilização de um algoritmo de agrupamento de dados, percebe-se que os tópicos estão mais homogêneos, tornando a busca de um leitor mais confortável.

Tabela 4 - Gêneros de livros agrupados

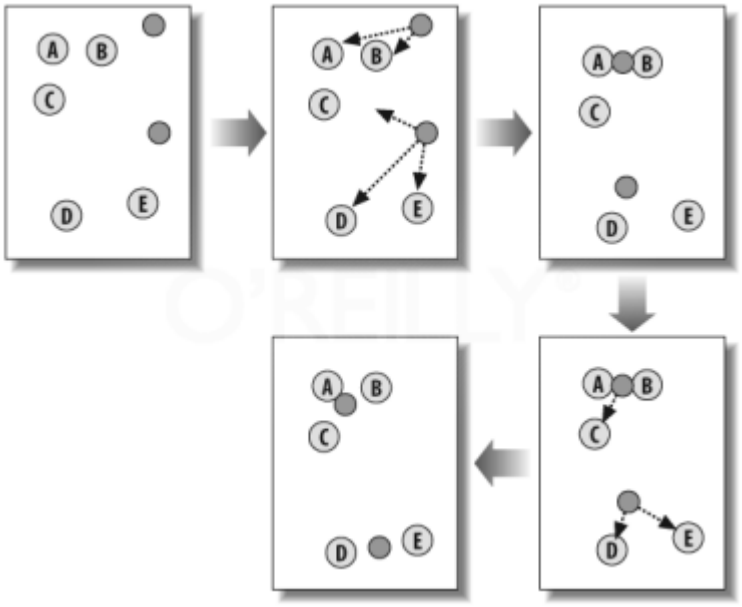
Prateleira 1	Religião
	História
	Geografia
Prateleira 2	Tecnologia
	Ficção Científica
	Esportes

Um dos algoritmos com mais uso para descoberta de conjuntos de dados é o K-Means, que tem como objetivo dividir uma quantidade de registros em um número *k* de grupos, minimizando a distância total dos dados de um grupo de dados e seu centro (PIMENTEL; FRANÇA; OMAR, 2003).

> "O número k de grupos que se deseja encontrar precisa ser informado de antemão. Em seguida, k pontos são escolhidos aleatoriamente para representar os centróides dos grupos, com isso, um conjunto de elementos, usualmente vetores, é particionado de forma que cada elemento é atribuído à partição, ou grupo, de centróide mais próximo, de acordo com a distância euclidiana comum. A cada iteração do algoritmo, os k centróides, ou "médias", e daí vem o nome means, são recalculados de acordo com os

elementos presentes no grupo e em seguida todos os elementos são realocados para a partição cujo o novo centróide se encontra mais próximo" (Lloyd 1982, apud COSTA et al., 2012).

Figura 3 - Funcionamento do Algoritmo K-Means



Fonte: Costa et al. (2012)

A Figura 3 demonstra o funcionamento passo-a-passo do algoritmo K-Means que De Castro e Do Prado (2001) apontam vantagens quando se refere a simplicidade e eficiência, pois tem rápido processamento em cálculos não complexos, permitindo trabalhar os dados de forma sequencial, sem a necessidade de grandes ocupações de memória.

2.1.3 Associação

A tarefa de associação tem foco na busca de relacionamento entre atributos, analisando e criando relacionamento entre eles, gerando o que é chamado de "regra de associação", dividindo a relação em antecedente e consequente. Amo (2004) afirma que as regras possuem um padrão de exibição X → Y, sendo que X e Y representam uma união de valores.

"É uma das tarefas mais conhecidas devido aos bons resultados obtidos, principalmente nas análises da "Cesta de Compras" (*Market Basket*), onde identificamos quais produtos são levados juntos pelos consumidores" (CAMILO; SILVA, 2009, p. 10).

Tabela 5 - Exemplos de regra de associação

Antecedente (X)	Consequente (Y)	Suporte	Confiança
{Fralda, Leite}	Cerveja	58%	80%
Notebook, Mochila	*Mouse*	70%	73%

A Tabela 5 apresenta duas regras de associação em seu formato mais comum de apresentação. Tomando como exemplo a segunda linha, a regra diz que em 70% das transações

(suporte) em que é comprado um *notebook* e mochila (antecedente), também é comprado um *mouse* (consequente), isso com uma confiança de 73%. Suporte é definido pela quantidade de vezes em que X e Y aparecem juntos no meio das transações. A confiança é a quantidade de vezes em que X e Y aparecem junto nas transações dividido pela quantidade de transações que contém somente os itens de X.

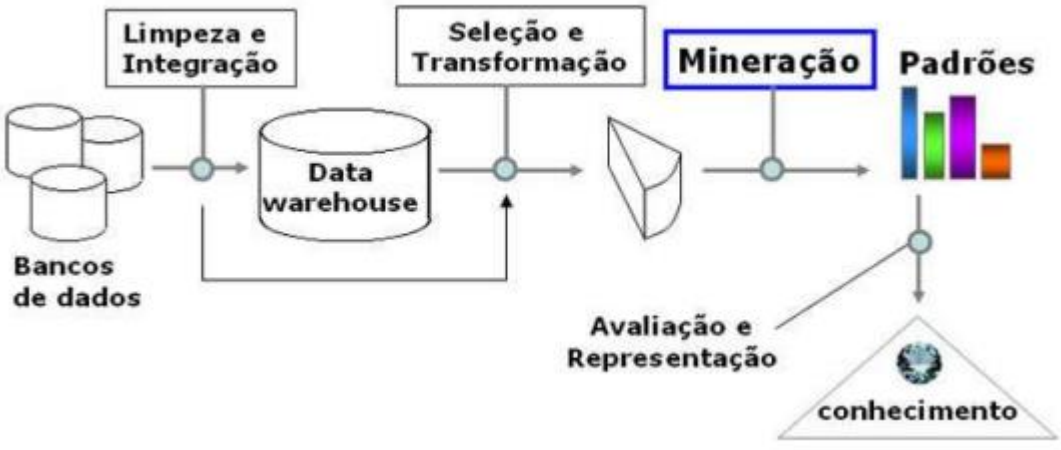
A associação também pode ser utilizada em exemplo como prever efeitos colaterais de um medicamento recém lançado (CAMILO; SILVA, 2009, p.10), ou também fazer a relação entre publicações de trabalhos científicos com as vezes em que um aluno de doutorado é orientado (CARDOSO; MACHADO, 2008, p. 506).

Segundo Han, Kamber e Pei (2012) os resultados da mineração de dados feita com a tarefas de associação podem servir de apoio para desenvolver métodos e estratégias de vendas e verificação de perfis de clientes. Em um setor de empréstimos bancários, a associação seria útil para avaliar os perfis de clientes que mais devem ou que melhor pagam os empréstimos realizados na empresa.

2.1.4 Processo KDD

Amo (2004) cita que muitas pessoas tratam mineração de dados como sendo o mesmo que Descoberta de Conhecimento em Banco de Dados ou *Knowledge Discovery in Databases* (KDD), mas na prática, o KDD é um processo maior, dividido nas etapas explicadas abaixo e ilustradas na Figura 4:

Figura 4 - Ilustração do processo KDD



Fonte: Amo (2004)

1. Limpeza dos dados: nesta etapa os dados são tratados de forma eliminar informações desnecessárias e inconsistentes;

2. Integração dos dados: os dados de diferentes origens são unidos de modo a formar uma única base de dados;
3. Seleção: etapa onde são selecionados os atributos que tenham relevância para o usuário. Por exemplo: o usuário pode decidir quais informações, como por exemplo a cor da pele e dos olhos, não são importantes para determinar a chance de uma pessoa ter problemas musculares;
4. Transformação dos dados: etapa na qual os dados são padronizados e colocados em formatos apropriados para que se possa aplicar os algoritmos de mineração de dados;
5. Mineração: importante etapa onde emprega-se de técnicas inteligentes para extração de padrões de interesse;
6. Avaliação: etapa em que identifica-se os padrões de interesse do usuário baseado em seus critérios;
7. Visualização dos resultados: nesta etapa usufrui-se de técnicas de representação do conhecimento a fim de apresentar ao usuário o conhecimento obtido através da mineração.

2.2 CRISP-DM

O CRISP-DM (*Cross-Industry Standard Process for Data Mining*) é um dos vários processos que definem e servem como base para o desenvolvimento de atividades de mineração de dados, isso explica-se pelo fato da existência vasta de literatura e documentações a respeito.

IBM Business Intelligence (2017) diz que o modelo de ciclo de vida do processo está dividido em seis etapas em um formato cíclico, sendo que a sequência do ciclo não precisa ser seguida exatamente como está definida, uma vez que os projetos tendem a irem e voltarem entre as fases, conforme faz-se necessário.

Em formato de metodologia, o processo, contém fases comuns e necessárias para projetos, as tarefas e o detalhamento de como as fases estão ligadas umas às outras. Quando visto como um modelo de referência, o CRISP-DM permite ter como base uma visão geral do ciclo de vida de um trabalho de mineração de dados (IBM BUSINESS INTELLIGENCE, 2017).

> "O modelo CRISP-DM é flexível e pode ser facilmente customizado. Por exemplo, se sua organização planejar detectar a lavagem de dinheiro, é provável que você examine detalhadamente grandes quantidades de dados sem uma meta de modelagem específica. Em vez da modelagem, seu trabalho irá se concentrar na exploração e visualização de dados para descobrir os padrões suspeitos em dados financeiros. O CRISP-DM permite que você crie um modelo de mineração de dados que se encaixe em suas necessidades específicas" (IBM BUSINESS INTELLIGENCE, 2017).

2.2.3 A Metodologia do CRISP-DM

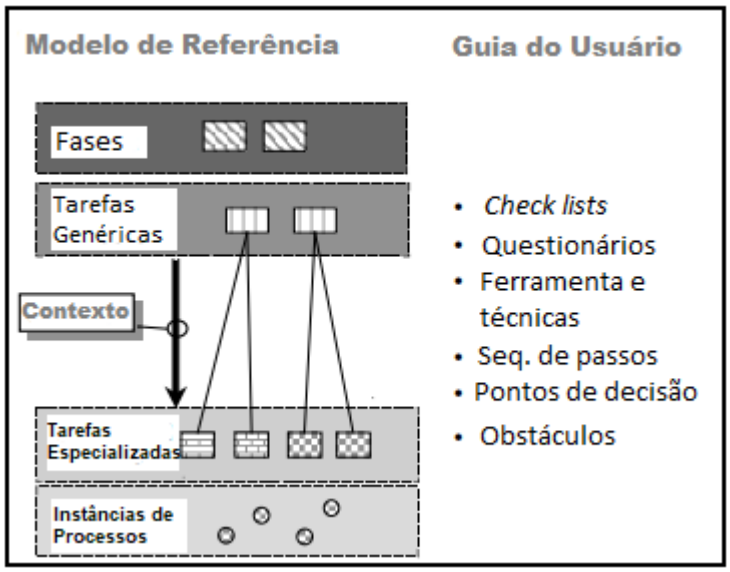
Seguindo a afirmação de Wirth e Hipp (2000), a metodologia do CRISP-DM está descrita em um processo de modelo hierárquico, dividido em quatro níveis de abstração, seguindo uma sequência do mais genérico para o mais específico: fases, tarefas genéricas, tarefas especializadas e instâncias de processos.

Wirth e Hipp (2000) também afirmam que o primeiro nível trata-se de um pequeno número de fases, onde em cada uma delas, existem vários segundos níveis de tarefas genéricas. Estas tarefas são consideradas genéricas por serem abrangentes o suficiente para cobrirem todas as situações possíveis de mineração de dados.

As tarefas especializadas, que são o terceiro nível da hierarquia da metodologia, é onde deve-se descrever como as ações das tarefas genéricas deverão ser executadas em situações específicas. Por exemplo, no segundo nível existe uma tarefa genérica chamada *modelo de construção*, o papel do terceiro nível é, existir uma tarefa *executar modelo de construção* que contém as atividades específicas para a solução do problema e quais ferramentas utilizar para tal.

O último nível, instância de processos, é um registro das ações, decisões e resultados de uma atual execução de mineração de dados. Esta pode ser organizada seguindo a ordem utilizada nas tarefas dos níveis anteriores, mas representa o que ocorreu em uma execução específica, ao invés de algo mais geral.

Figura 5 - Os quatro níveis da metodologia CRISP-DM



Fonte: Traduzido de Wirth e Hipp (2000)

A Figura 5 ilustra os níveis da metodologia do CRISP-DM, apresentando como ocorre a hierarquia entre as fases e um Guia do Usuário com dicas e sugestões de materiais que estão envolvidos em um projeto de mineração de dados.

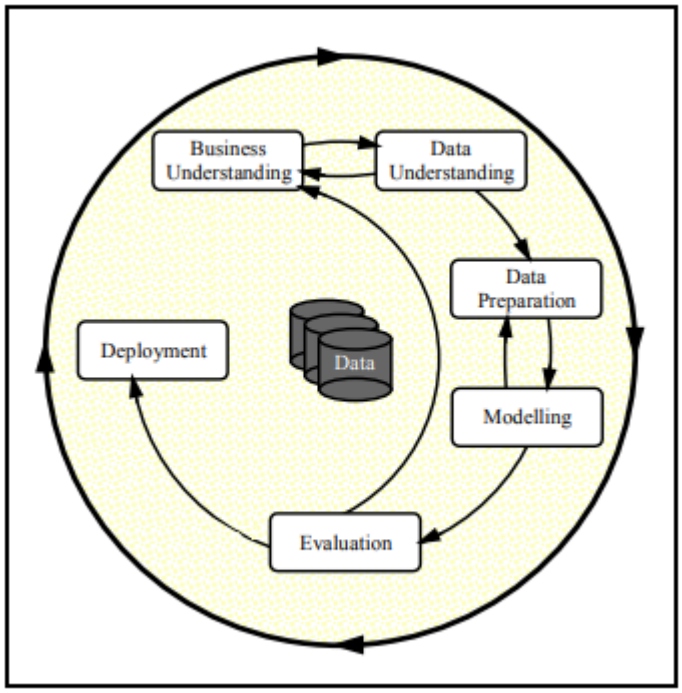
2.2.4 Modelo de Referência CRISP-DM

O modelo de referência do CRISP-DM oferece uma visão geral do ciclo de vida de um projeto de mineração de dados (WIRTH; HIPP, 2000). Em Olson e Delen (2008, apud Camilo e Silva, 2009), o modelo de referência divide-se em seis partes que estão dispostas de maneira cíclica, porém, a sequência das etapas não é rigorosa, permitindo assim que a execução de um projeto de mineração de dados possa ir e vir entre as fases existentes.

É possível verificar na

Figura 6 que, além do ciclo das seis fases do CRISP-DM, existe um ciclo externo que envolve todo o processo, Wirth e Hipp (2000) afirmam que este ciclo simboliza o processo natural da mineração de dados, sendo que, a mineração de dados não é finalizada após a implantação de uma solução, pois as lições aprendidas em projetos anteriores podem gerar novas questões e modelos de negócio.

Figura 6 - Ciclo de vida do processo CRISP-DM



Fonte: Wirth e Hipp (2000)

Abaixo estão descritas as características de cada fase do modelo de referência do CRISP-DM.

- Entendimento do negócio (*Business Understanding*): Camilo e Silva (2009) dizem que esta fase tem foco entender qual o objetivo deseja-se atingir com a mineração de

dados. Pode ser algo não tão fácil de executar quanto parece, sendo que é possível diminuir os riscos futuros através de metas e entendimento de problemas (IBM BUSINESS INTELLIGENCE, 2017).

- Entendimento dos dados (*Data Understanding*): esta fase começa com um conjunto inicial de dados e procedimento que ajudarão na familiarização desses dados (AZEVEDOS; SANTOS, 2008). Como dito por Olson e Delen (2008, apud Camilo e Silva, 2009), é necessário, após a definição dos objetivos o conhecimento dos dados visando uma boa descrição do problema, levantar quais dados são importantes para a solução e garantir que as variáveis importantes para o projeto não sejam interdependentes.
- Preparação dos dados (*Data preparation*): considerando que os dados podem ter diversas origens, pode acontecer que os mesmos não estejam preparados em um formato que seja possível realizar a mineração, e dependendo da situação em que estes se encontram, torna-se necessário utilizar alguns métodos para filtrar, limpar e preencher valores vazios nestes dados (CAMILO; SILVA, 2009). Segundo IBM Business Intelligence (2017), esta é uma das fases que mais leva tempo em um projeto de mineração de dados, consumindo de 50 a 70% do tempo de todo o projeto, tempo que pode ser reduzido ou não estendido se as fases anteriores ocorrerem de maneira adequadas.
- Modelagem (*Modelling*): fase onde ocorre a aplicação das técnicas e tarefas de mineração de dados, acontecendo geralmente em várias iterações. IBM Business Intelligence (2017) destaca que "normalmente, os mineradores de dados executam diversos modelos usando os parâmetros padrão e, então, ajustam os parâmetros ou revertem para a fase de preparação de dados para as manipulações requeridas por seu modelo de preferência".
- Avaliação (*Evaluation*): etapa que exige a presença dos especialistas dos dados e entendedores do negócio, uma vez que são feitos testes e avaliações para que se possa obter confiança nos modelos obtidos. Estas atividades são necessárias considerando que esta é uma das fases mais críticas do modelo de referência (CAMILO; SILVA, 2009)
- Implantação (*Deployment*): fase na qual os dados são apresentados às partes interessadas. Não se trata necessariamente da etapa final do projeto, pois uma vez que o cliente conhece os resultados gerados, isto pode gerar novos requisitos e demandas que tornarão necessárias novas implementações (WIRTH; HIPP, 2000).

2.3 TRABALHOS RELACIONADOS

Esta seção apresenta trabalhos de mineração de dados voltados para a análise de dados educacionais.

2.3.1 Técnicas de mineração de dados aplicadas aos microdados do ENADE para avaliar o desempenho dos acadêmicos do curso de Ciência da Computação do Rio Grande do Sul utilizando o software R.

Com o uso de *Hierarchical Clustering* (Agrupamento Hierárquico) e da linguagem de programação R, Vista, Figueiró e Chicon (2017) realizaram a mineração dos microdados do ENADE para avaliação do desempenho dos acadêmicos dos cursos de Ciência da Computação no estado do Rio Grande do Sul. Os resultados alcançados possibilitaram a identificação do nível de desempenho acadêmicos de cada instituição do Estado, promovendo o apoio na tomada de decisão no que tange a melhoria do ensino superior brasileiro.

Os primeiros resultados obtidos foram o levantamento dos dados das variáveis do ENADE 2014, que apresentava informações como número mínimo e máximo de inscritos para realização da prova, número de participantes e o Conceito ENADE das universidades. A partir disto, foi realizada a tarefa de agrupamento nos dados, atividade esta que resultou em quatro grupos de instituições de ensino superior.

- Grupo 1: composto por nove instituições, onde seis obtiveram Conceito ENADE 3 e outras três obtiveram Conceito ENADE 4. O Conceito ENADE Contínuo apresentou 2,86 de média.
- Grupo 2: com dez instituições no total, das quais sete obtiveram Conceito ENADE 2 e somente duas obtiveram Conceito 3. A média do Conceito ENADE Contínuo foi de 1,74.
- Grupo 3: apenas uma instituição com Conceito ENADE 1 e Conceito ENADE Contínuo de 0,43.
- Grupo 4: três instituições, onde duas obtiveram Conceito ENADE 4 e apenas uma obteve Conceito ENADE 5. A média do Conceito ENADE Contínuo foi de 5,53.

Com os grupos acima pode-se observar que as instituições contidas no grupo quatro foram as que obtiveram melhores resultados no ENADE no ano de 2014, pois seus Conceitos ENADE Contínuo foram os mais altos em uma escala de 0 a 5. As instituições do grupo 1 ficaram em segundo lugar, também apresentando bons resultados. A instituição presente no grupo 1 ficou isolada em seu *cluster*, pois foi a que apresentou pior resultado.

2.3.2 Prática de Mineração de Dados no Exame Nacional do Ensino Médio

Utilizando da tarefa de associação, Silva, Morino e Sato (2014) utilizaram o algoritmo de Apriori na tentativa de encontrar padrões de associação nos resultados de provas e questionários socioeconômicos do Exame Nacional do Ensino Médio (ENEM). Os dados utilizados para o desenvolvimento do trabalho foram coletados diretamente do site do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), que os dispõe de forma aberta.

A primeira etapa definiu a região do país que seria foco da pesquisa, onde foram selecionados dados somente das capitais da região Sudeste do país, como um total de 452.710 alunos, fez-se necessário também a eliminação de registros dos alunos que não compareceram nos dois dias de prova, um total aproximado de 310.000.

As perguntas selecionadas para a pesquisa foram voltadas para a quantidade de membros na família do estudante, escolaridade da mãe, renda familiar e em que tipo de escola o aluno estudou durante o Ensino Médio. Tais perguntas foram escolhidas com o intuito de saber se, de acordo com as informações, há influência na nota e no desempenho do aluno na prova do ENEM.

Os resultados obtidos apontaram que os fatores: renda familiar baixa, escolaridade em nível primário dos pais e um alto número de pessoas morando com o estudante, afetam negativamente o desempenho do aluno. Os autores da pesquisa concluíram que o ensino público no Brasil precisa de melhorias, tanto política quanto pedagogicamente, onde a classe social mais baixa é afetada, o que exerce influência direta no desempenho do estudante.

3 METODOLOGIA

Esta seção apresenta a metodologia aplicada para o desenvolvimento deste trabalho, citando materiais e processos utilizados.

3.1 OBJETO DE ESTUDO

Este trabalho teve por objetivo desenvolver um módulo da plataforma para visualização de resultados, gerados a partir da aplicação de técnicas e recursos de mineração de dados, em cima dos dados de respostas dos estudantes do curso de graduação em Psicologia do ENADE. Os dados que foram utilizados para estudo foram extraídos dos microdados disponibilizados pelo Inep, que estão dispostos de forma aberta em arquivos de formatos .csv e contêm informações dos acadêmicos, respostas das provas e os gabaritos de cada área cujo as provas foram realizadas.

3.2 MATERIAIS

A seguir, na seção 3.2.1, estão apresentadas as tecnologias e ferramentas que serviram de apoio para o completo desenvolvimento do trabalho.

3.2.1 Tecnologias e ferramentas

Para atender as necessidades de tratamento, limpeza, padronização, mineração e armazenamento dos dados que foram coletados, foram utilizadas as seguintes ferramentas computacionais:

- Microsoft Excel: editor de planilhas que atende diversos formatos de arquivos, dentre eles, o formato .csv, que corresponde ao formato dos dados disponibilizados pelo ENADE. Esta ferramenta serviu de apoio no tratamento e limpeza dos dados;
- Linguagem Python: Python é uma linguagem de programação de alto nível e serviu, em conjunto com o Microsoft Excel, de apoio para o tratamento dos dados, de forma a permitir a construção de códigos para lidar com grandes quantidades de dados;
- Linguagem R e R Studio: a linguagem de programação R, em conjunto com seu ambiente integrado R Studio, permitiram a descoberta de novos conhecimentos através da mineração dos dados. Estas ferramentas possuem bibliotecas com diversos tipos de algoritmos que possibilitaram atender os objetivos da proposta;
- SQLite: ferramenta que permite a utilização de recursos de banco de dados sem a necessidade de utilização de um SGBD. Sua utilização é simplificada, sendo necessário somente um arquivo para armazenamento e leitura dos dados.

- Weka (Waikato Environment for Knowledge Analysis): esta ferramenta agrupa algoritmos das diferentes abordagens de mineração de dados e outras atividades como Inteligência Artificial. A mesma permitiu a busca de resultados com a tarefa de agrupamento.
- Pentaho: software de código aberto que permitiu, após os tratamentos dos dados, fazer a carga dos mesmos na base de dados utilizada para armazenamento dos resultados encontrados no trabalho.

Para apresentação dos resultados obtidos, foi criada uma página Web desenvolvida com tecnologias de programação *front-end*. Para tanto, foram utilizadas as seguintes tecnologias:

- Visual Studio Code: editor de texto completo, com diversos recursos que permitem a rápida edição de código-fonte de softwares;
- Framework Angular: plataforma da Google que permite a criação de aplicações *front-end* escritas com base na linguagem de programação TypeScript;
- Framework CSS Bootstrap 4: conjunto de recursos prontos para criação de páginas Web com visuais amigáveis e estilizados.

3.3 PROCEDIMENTOS

As respostas analisadas durante o projeto foram, especificamente, dos acadêmicos do curso de Psicologia das instituições de ensino superior do Brasil.

A partir do modelo de referência CRISP-DM, foram realizadas as adaptações necessárias para o desenvolvimento deste trabalho. A

Figura 7 ilustra as etapas da metodologia que foram realizadas durante o trabalho, sendo que as mesmas são explanadas em tópicos posteriormente.

Figura 7 - Etapas do desenvolvimento da proposta

Nota-se, na

Figura 7, a divisão das nove etapas que dividem o desenvolvimento do trabalho. Para melhor entendê-las, a seguir estão os detalhamentos de cada uma das divisões do processo metodológico.

- **Definição da proposta:** a etapa inicial deste trabalho consistiu em, juntamente à orientadora do projeto, estabelecer a proposta a ser desenvolvida. A definição da proposta envolveu escolher a área a ser estudada e quais tipos de resultados poderiam ser esperados da análise dos dados, já tentando também identificar uma forma viável de apresentar os resultados.

  Decidiu-se que as perguntas contidas nas provas do ENADE deveriam ser distribuídas com base nas áreas de estudos do curso de Psicologia do CEULP/ULBRA. Esta divisão foi baseada na matriz curricular 04388 - passada pela coordenação do curso de Psicologia da instituição através de uma entrevista de melhor conhecimento da área - que divide as matérias em seis grandes áreas. A

Tabela 6 apresenta a forma como as áreas estão divididas.

Tabela 6 – Estrutura curricular do curso de Psicologia do CEULP/ULBRA (Matriz 043888)

Eixos	**Matérias**
Fundamentos epistemológicos e	● História e Sistemas da Psicologia ● Ética e Legislação em Psicologia

históricos	• Filosofia • Antropologia
Fundamentos teórico-metodológicos	• Fundamentos das Medidas Psicológicas • Técnicas de Entrevista Psicológica • Métodos e Técnicas de Avaliação Psicológica I • Métodos e Técnicas de Avaliação Psicológica II • Teorias e Técnicas de Dinâmica de Grupo • Teorias e Técnicas psicoterápicas I • Teorias e Técnicas psicoterápicas II • Teorias e Técnicas psicoterápicas III • Teorias e Técnicas psicoterápicas IV
Procedimentos para a investigação científica e a prática profissional	• Instrumentalização Científica • Saúde Mental e Trabalho • Avaliação Neuropsicológica • Pesquisa em Psicologia • TCC I • TCC II
Fenômenos e processos psicológicos	• Processos Básicos em Psicologia • Psicologia do Desenvolvimento I • Psicologia do Desenvolvimento II • Psicologia da Personalidade • Psicologia Social • Psicologia Experimental • Psicologias da Aprendizagem • Psicologia das Relações Familiares • Psicologia Comunitária • Psicopatologia Geral I • Psicopatologia Geral II • Psicologia da Saúde • Psicologia da Educação • Psicologia do Trabalho • Psicologia nas Organizações
	• Comunicação e Expressão • Morfofisiologia e comportamento humano

Interfaces com campos afins do conhecimento	● Bases biológicas do comportamento humano ● Saúde, Bioética e Sociedade ● Psicofarmacologia ● Estatística Aplicada à Psicologia ● Neuropsicologia ● Sociedade e Contemporaneidade ● Tópicos Especiais em Psicologia ● Cultura Religiosa
Práticas profissionais	● Estágio Básico I ● Estágio Básico II ● Estágio Básico III ● Estágio Básico IV ● Estágio Básico V ● Intervenção em Grupos ● Intervenções em Situações de Crise ● Estágio Específico na Ênfase I (A) ● Estágio Específico na Ênfase I (B) ● Estágio Específico na Ênfase II (A) ● Estágio Específico na Ênfase II (B)

Fonte: PPC Psicologia, 2018

- **Extração dos dados:** nesta etapa foram levantadas as fontes das quais os dados seriam extraídos e quais dados seriam analisados. De acordo com a proposta, os dados a serem extraídos e analisados foram as respostas dos estudantes da área de Psicologia correspondente aos anos definidos para estudo, analisando também os gabaritos das provas, para que seja possível o estudo da quantidade de erros e acertos dos discentes que realizaram as provas. Os dados de todas as provas realizadas pelo ENADE estão disponíveis abertamente na página de microdados do Inep.
- **Entendimento dos dados:** com os dados extraídos, torna-se necessário o entendimento da estrutura e da forma como estão organizados, bem como o conhecimento das informações contidas nos mesmos com apoio do dicionário de variáveis que é disponibilizado pelo próprio ENADE juntamente com os dados.
- **Adequações da proposta:** foram feitas adequações na proposta à medida com que o conhecimento sobre os dados foi adquirido, de forma a garantir que não fossem realizadas tarefas que fogem do proposto e que estudos considerados importantes façam parte do projeto. Dentre as adequações, encontra-se a redução da

-
-

- Tabela 6, pois existem matérias da estrutura curricular que não possuíam questões nas provas do ENADE.
- **Adequação dos dados para análise:** com as adequações garantidas anteriormente, os dados extraídos foram transformados e padronizados de modo que os algoritmos de mineração de dados utilizados pudessem trabalhar corretamente. Para dinamização deste processo, foram utilizadas ferramentas como Microsoft Excel e Pentaho ETL. As ferramentas também serviram de auxílio na remoção de inconsistências presentes nos dados das provas e dos gabaritos respectivos.
- **Aplicação dos algoritmos de mineração de dados:** os dados, nesta etapa, já se encontravam prontos para serem analisados pelos algoritmos selecionados para a mineração. A aplicação dos algoritmos foi realizada através da utilização de recursos da linguagem de programação R, em conjunto com sua plataforma de desenvolvimento R Studio. A ferramenta Weka também serviu de apoio na análise dos dados durante a utilização da tarefa de agrupamento. Tais ferramentas possuem recursos para leituras de arquivos .csv e de bases de dados, e suas saídas podem ser dadas de diferentes formas, dependendo do algoritmo que está sendo executado.
- **Análise e validação dos resultados:** os resultados obtidos através da mineração foram avaliados, com o intuito de garantir que estes atendam os objetivos propostos para o projeto. Foram avaliadas as respostas e taxas de acerto dos estudantes de Psicologia, fazendo levantamento das áreas mais eficientes e deficientes dos discentes, e também dos perfis de estudantes das instituições mais bem conceituadas no ENADE, podendo servir como ferramenta de apoio à decisão. As análises foram realizadas juntamente com o acompanhamento da orientadora do projeto e da especialista da área de Psicologia do CEULP/ULBRA. A participação deste especialista serviu para a validação dos resultados e obtenção de novas informações relevantes sobre a análise dos dados.
- **Elaboração da plataforma para apresentação dos resultados:** a plataforma foi desenvolvida com o uso dos recursos computacionais citados na seção 3.2.1. Esta plataforma serve para apresentação intuitiva dos resultados obtidos através da mineração de dados realizada. A principal forma de apresentação dos dados é feita por gráficos, de maneira que facilitem a interpretação por parte do público. A visualização

tem diferentes pontos de vista, tornados possíveis através da elaboração de filtros dinâmicos.

- **Disponibilização da plataforma:** após a conclusão de todas as etapas descritas anteriormente, a plataforma foi disponibilizada, de forma a estar acessível para o público que possui interesse na visualização dos tipos de informações dispostas.

4 RESULTADOS E DISCUSSÃO

Esta seção apresentará os passos de desenvolvimento do trabalho, passos esses que foram seguidos de acordo com a metodologia proposta no trabalho. Os resultados estão divididos de acordo com os tipos de resultados que foram encontrados durante o trabalho. Na seção 4.1 encontra-se a descrição dos resultados obtidos com a tarefa de agrupamento e, na seção 4.2, os resultados alcançados com a tarefa de associação.

4.1 RESULTADOS DE AGRUPAMENTO

Aqui estão presentes os resultados obtidos com a utilização da tarefa de agrupamento. Tais resultados mostram o agrupamento das questões respondidas pelos estudantes divididas de acordo com os conteúdos curriculares do curso de Psicologia do CEULP/ULBRA.

4.1.1 Extração dos dados

Os dados a serem trabalhados foram coletados no site do Inep, plataforma oficial para disponibilização dos microdados do ENADE. Tais dados são disponibilizados em tabelas no formato .csv. Seguindo a proposta do trabalho, foram coletados os dados somente dos anos de 2009, 2012 e 2015, anos esses que correspondem aos acontecimentos das provas dos estudantes da área de Psicologia.

4.1.2 Entendimento dos Dados

Após a baixar os dados dos anos correspondentes ao curso de Psicologia, os mesmos foram, junto aos seus respectivos dicionários de variáveis, observados com objetivo de definir quais colunas seriam utilizadas para continuar o desenvolvimento proposto no trabalho. As colunas mantidas para o desenvolvimento do trabalho para a parte de agrupamento foram

- NU_ANO: que representa o ano de prova dos dados;
- CD_CATAD: indica a categoria administrativa da IES;
- CO_GRUPO: código de referência ao curso do estudante;
- CO_REGIAO_HABIL: representa a região de funcionamento da instituição;
- CO_UF_HABIL: representa a unidade federativa da instituição;
- TP_PR_OB_CE: indica o tipo de presença na parte objetiva do componente específico;
- VT_GAB_OCE: gabarito da parte objetivado componente específico. A coluna foi renomeada para GABARITO_ESPECIFICO;
- VT_ESC_OCE: indica as respostas do estudante na parte objetiva do componente específico. A coluna foi renomeada para RESPOSTAS_ESPECIFICO;

Figura 8 - Colunas selecionadas para o ano de 2009

nu_ano	cd_catad	co_grupo	co_regiao_habil	co_uf_habil	tp_pr_ob_ce	gabarito_especifico	respostas_especifico
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CCDDCCBBCDABEABEAEABCAABCAE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CEBAEACEABADBEECADBEEDBCDBB
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ABBCCCAADBAABEBEADCBCADDCCE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ECCEDCADABAAEDBCAADDCADECAE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ACBDCCAAAEADEDEEADCBEAAAAAB
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	EBCDDAAEBEABEEBEAABBEAACBCE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CBEEABECACABECDCAADECAACCAE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ECDEEBECBCACDEADABDBDDDAAEE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CCBDACBCABAAEAAEAAABBADCAAE

A Figura 8 apresenta a organização das colunas selecionadas no desenvolvimento do trabalho para um dos três anos que foram minerados.

4.1.3 Adequações da Proposta

Durante o entendimento dos dados e enquadramento das questões foi possível perceber que nem todos os conteúdos curriculares possuíram questões, e os conteúdos que tiveram questões que se enquadraram variaram de acordo com o ano de prova. Portanto, cada ano de dado minerado possui resultados de conteúdos curriculares que não estão presentes em outros anos. Por exemplo, no ano de 2009, o conteúdo de Antropologia teve duas questões enquadradas, porém, na prova de 2012, este conteúdo não recebeu questões relacionadas.

Tabela 7 - Enquadramento das questões nos conteúdos curriculares para o ano de 2015

Eixo curricular conforme diretrizes	Conteúdo Curricular	Questão
I - Fundamentos epistemológicos e históricos	História e Sistemas da Psicologia	9, 11, 21
	Ética e Legislação em Psicologia	
	Filosofia	
	Antropologia	
II - Fundamentos teórico-metodológicos	Fundamentos das Medidas Psicológicas	12
	Técnicas de Entrevista Psicológica	
	Métodos e Técnicas de Avaliação Psicológica	17, 31
	Teorias e Técnicas de Dinâmica de Grupo	22
	Teorias e Técnicas Psicoterápicas - Psicanálise	
	Teorias e Técnicas Psicoterápicas - Comportamental	
	Teorias e Técnicas Psicoterápicas - Humanismo	
	Teorias e Técnicas Psicoterápicas - Sistêmica	

III - Procedimentos para a investigação científica e a prática profissional	Instrumentalização Científica	
	Saúde Mental e Trabalho	
	Avaliação Neuropsicológica	
	Pesquisa em Psicologia	14, 15, 16
	Intervenção em Situações de Crise	
	TCC	
IV - Fenômenos e processos psicológicos	Processos Básicos em Psicologia	27
	Psicologia do Desenvolvimento	13, 20
	Psicologia da Personalidade	23
	Psicologia Social	
	Psicologia Comunitária	
	Psicologias da Aprendizagem	
	Psicologia Experimental	19, 24
	Psicologia das Relações Familiares	
	Psicopatologia	18
	Psicologia da Saúde	33
	Psicologia da Educação	28
	Psicologia do Trabalho	30
	Psicologia nas Organizações	29
V - Interfaces com campos afins do conhecimento	Morfofisiologia e comportamento humano	
	Bases biológicas do comportamento humano	34
	Saúde, Bioética e Sociedade	32
	Psicofarmacologia	
	Estatística	
	Neuropsicologia	25, 26
	Sociedade e Contemporaneidade	
	Psicologia da Comunicação	
	Psicologia Jurídica	
	Psicologia da Sexualidade Humana	
	Psicossomática	
	Psicologia do Esporte	
	Psicologia Hospitalar	
	Psicologia Positiva	
VI - Práticas profissionais	Intervenção em Grupos	35

	Estágios Básicos	10
	Estágios em Ênfases	

A Tabela 7 mostra o enquadramento das questões do ano de 2015 seguindo a estrutura curricular definida na proposta de trabalho. Nota-se que nem todos os conteúdos curriculares possuem questões enquadradas e alguns deles possuem mais questões do que outros.

4.1.4 Adequações dos Dados

A primeira etapa de trabalho com os dados consistiu em separar respostas somente dos alunos da área de Psicologia para os três anos de prova. Considerando que o desenvolvimento deste trabalho limitou-se somente aos dados dos estudantes da área de Psicologia, fez-se necessário exclusão dos dados de estudantes de outras áreas. De acordo com o dicionário de variáveis dos dados, o código de grupo da área de Psicologia correspondia ao 18, portanto, através de filtro da ferramenta Microsoft Excel, foram excluídas todas as linhas nas quais a coluna `CO_GRUPO` não possuía o número 18. O procedimento foi repetido para os três anos de dados.

Após, foram separados os dados somente dos alunos que tiveram presença total na parte objetiva do componente específico, isso foi possível mantendo na planilha apenas as linhas cuja coluna `TP_PR_OB_CE` possuía o código 555, código este que indica a presença total.

Por seguinte, foram mantidas apenas colunas relevantes para a descoberta de conhecimento do trabalho, evitando também a presença de dados inconsistentes no momento da utilização dos algoritmos. As colunas mantidas estão descritas na seção 4.2.

Figura 9 - Estado dos dados após primeiras adequações

nu_ano	cd_catad	co_grupo	co_regiao_habil	co_uf_habil	tp_pr_ob_ce	gabarito_especifico	respostas_especifico
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CCDDCCBBCDABEABEAEABCAABCAE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CEBAEACEABADBEECADBEEDBCDBB
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ABBCCCAADBAABEBEADCBCADDCCE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ECCEDCADABAAEDBCAADDCADECAE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ACBDCCAAAEADEDEEADCBEAAAAAB
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	EBCDDAAEBEABEEBEAABBEAACBCE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CBEEABECACABECDCAADECAACCAE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ECDEEBECBCACDEADABDBDDDAAEE
2009	1	18	5	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CCBDACBCABAAEAAEAAABBADCAAE

Com a execução dos passos anteriores, os dados encontravam-se no estado como mostrado na Figura 9. Com isso, foi criado um *script* para comparação das respostas de cada estudante com as repostas do gabarito final.

Figura 10 - *Script* de adição de colunas de questões

```
for i in range(len(list_of_answers[1])):
    dataset['questao_' + str(i + 1)] = ''
```

O código da Figura 10 adiciona colunas à planilha com os títulos `questao_1`, `questao_2`, ..., `questao_x`. A quantidade de colunas adicionadas segue o número de questões do gabarito final das provas, para os três anos de provas considerados neste trabalho, o número final de questões foi um total de 27.

Figura 11 - Código de comparação das respostas com o gabarito final

```
for i in range(students_answers.count()):
    if dataset['co_regiao_habil'][i] == 1:
        dataset['co_regiao_habil'][i] = 'Norte'
    elif dataset['co_regiao_habil'][i] == 2:
        dataset['co_regiao_habil'][i] = 'Nordeste'
    elif dataset['co_regiao_habil'][i] == 3:
        dataset['co_regiao_habil'][i] = 'Sudeste'
    elif dataset['co_regiao_habil'][i] == 4:
        dataset['co_regiao_habil'][i] = 'Sul'
    elif dataset['co_regiao_habil'][i] == 5:
        dataset['co_regiao_habil'][i] = 'Centro-Oeste'

    print('Comparando linha: ', i + 1)
    for j in range(len(students_answers[i])):
        if list_of_answers[i][j] == '8':
            dataset['questao_' + str(j + 1)][i] = 'X'
        else:
            if students_answers[i][j] == '*' or students_answers[i][j] == '_':
                dataset['questao_' + str(j + 1)][i] = 'X'
            else:
                if students_answers[i][j] == list_of_answers[i][j]:
                    dataset['questao_' + str(j + 1)][i] = 'Certa'
                else:
                    dataset['questao_' + str(j + 1)][i] = 'Errada'
```

Na Figura 11, o código mostrado percorre todas as linhas (dados de cada estudante) da tabela fazendo um tratamento de dados, primeiramente, ao entrar em uma linha, o algoritmo verifica se na coluna `CO_REGIAO_HABIL` o código é 1, 2, 3, 4 ou 5, e, dependendo do caso, faz uma substituição do dado por, Norte, Nordeste, Sudeste, Sul ou Centro-Oeste, respectivamente.

O segundo passo do algoritmo mostrado acima é comparar cada resposta de cada estudante com cada resposta do gabarito final e, caso a resposta do estudante para uma questão seja igual a resposta da mesma questão no gabarito final, a coluna daquela questão é marcada com o valor ‘Certa’, caso essa condição não seja satisfeita, é atribuído o valor ‘Errada’. Este passo também faz algumas validações: caso uma das respostas do gabarito final possua o valor ‘8’, a coluna da questão respectiva recebe o valor ‘X’, pois significa que aquela é uma questão anulada. Validações do tipo também são feitas nas respostas dos estudantes: caso alguma das

questões possua o valor ‘*’ ou ‘_’, a coluna da questão correspondente recebe o valor ‘X’, pois isto significa alguma irregularidade na resposta, como o estudante ter deixado em branco ou ter marcado mais de uma opção objetiva.

Figura 12 - Nova composição da tabela após execução do *script* de comparação de respostas

nu_ano	cd_catad	co_grupo	co_regiao_habil	co_uf_habil	tp_pr_ob_ce	gabarito_especifico	respostas_especifico	questao_1	questao_2	questao_3	questao_4	questao_5
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CCDDCCBBCDABEABEAEABCAABCAE	Certa	Certa	Errada	Certa	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CEBAEACEABADBEECADBEEDBCDBB	Certa	Errada	Certa	Errada	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ABBCCCAADBAABEBEADCBCADDCCE	Errada	Errada	Certa	Errada	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ECCEDCADABAAEDBCAADDCADECAE	Errada	Certa	Errada	Errada	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ACBDCCAAAEADEDEEADCBEAAAAAB	Errada	Certa	Certa	Certa	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	EBCDDAAEBEABEEBEAABBEAACBCE	Errada	Errada	Errada	Certa	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CBEEABECACABECDCAADECAACCAE	Certa	Errada	Errada	Errada	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	ECDEEBECBCACDEADABDBDDDAAEE	Errada	Certa	Errada	Errada	Errada
2009	Federal	18	Centro-Oeste	51	555	CCBDBECBAEACEEDDAA8BCADECAE	CCBDACBCABAAEAAEAAABBADCAAE	Certa	Certa	Certa	Certa	Errada

A Figura 12 é uma amostra da nova composição da tabela de respostas de estudantes após a execução do *script* de comparação. Esta etapa foi essencial para permitir que os algoritmos de mineração de dados pudessem ser utilizados para o restante do desenvolvimento do trabalho.

Após os resultados anteriores, notou-se que algumas regiões possuíam um número consideravelmente maior de estudantes do que outras, portanto, foi decidido separar cada região, para evitar inconsistência de resultados durante a mineração, o que poderia resultar que a mineração de uma região com muitos dados ofuscasse os resultados de uma outra com menor volume.

Tabela 8 - Quantidade de respostas de cada região para cada ano de prova

Ano	Total	Região	Número de respostas	Porcentagem
2009	40415	Norte	1897	4,69%
		Nordeste	7064	17,47%
		Centro-Oeste	3405	8,42%
		Sudeste	20165	49,89%
		Sul	7884	19,50%
2012	19889	Norte	1148	5,77%
		Nordeste	3943	19,82%
		Centro-Oeste	1517	7,62%

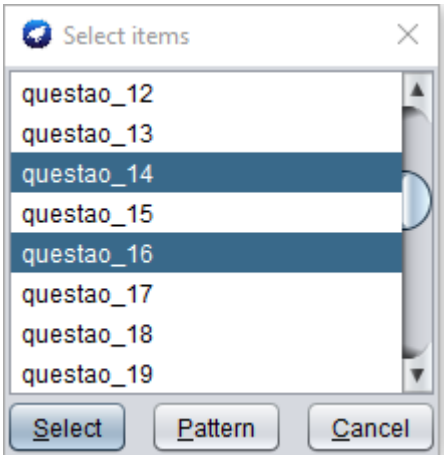
		Sudeste	9423	47,37%
		Sul	3858	19,39%
2015	24155	Norte	1258	5,20%
		Nordeste	5367	22,21%
		Centro-Oeste	2021	8,36%
		Sudeste	10934	45,26%
		Sul	4575	18,94%

A Tabela 8 apresenta a quantidade de respostas de cada região de cada ano de prova. É possível notar que a região Norte possui um número consideravelmente menor de respostas se for feita comparação com outras regiões. Percebe-se também que a região Sul é a que possui o maior volume de estudantes que fizeram a prova, isso, minerado junto à região Norte, causou inconsistência nos resultados da tarefa de agrupamento, apresentando saídas que poderiam não representar a realidade. O ano de 2009, comparado aos anos de 2012 e 2015, possui um número mais elevado de estudantes, justificado pelo fato de que, em 2009, além dos concluintes, também participavam do exame acadêmicos ingressantes nas IES.

4.1.5 Aplicação dos Algoritmos

Com os dados todos preparados e adequados de forma que a mineração de dados fosse possível, foram testados algoritmos de diferentes tarefas, porém, a tarefa que trouxe resultados mais consistentes foi a de agrupamento com o algoritmo Simple KMeans, algoritmo este que possui um bom desempenho no agrupamento de atributos, portanto, foi bastante eficaz na mineração dos dados na forma como estava estruturados. A utilização do algoritmo foi testada com a linguagem R, no entanto, quando utilizado nesta linguagem, o mesmo só aceita trabalhar com valores numéricos, portanto, foi decidido trabalhar com a ferramenta Weka, que já possui um tratamento automático para valores não numéricos.

Figura 13 - Filtro de atributos do Weka



A ferramenta também permite o filtro de atributos dos dados a serem minerados, o que facilita a mineração selecionando apenas a região e as questões desejadas das áreas de enquadramento, como é possível ver na Figura 13.

Figura 14 - Exemplo de agrupamento com o algoritmo Simple KMeans

```
Cluster 0: Sul,Certa,Errada
Cluster 1: Sul,Errada,Errada
Cluster 2: Sul,Errada,Certa
Cluster 3: Sul,Certa,Certa
Cluster 4: Sul,Certa,X
```

A Figura 14 apresenta um exemplo de como são realizados os agrupamentos pelo algoritmo Simple KMeans. Neste exemplo de mineração, foram selecionadas para processamento no algoritmo as questões 12 e 14 – que fazem parte do conteúdo curricular de História e Sistemas da Psicologia – e foi limitado também somente respostas de alunos da região Sul do Brasil.

Figura 15 - Exemplo de volume de incidências do algoritmo Simple KMeans

```
Clustered Instances

0       3479 ( 44%)
1       2547 ( 32%)
2        662 (  8%)
3       1097 ( 14%)
4         99 (  1%)
```

Na Figura 15 é possível ver o volume de incidências de cada grupo exibido na Figura 14, como exemplo: o *Cluster* 0, que representa um grupo de estudantes da região Sul que acertaram a questão 12 e erraram a questão 14, teve um volume de incidências de 3479 estudantes, o que corresponde a 44% do total da região Sul que responderam a prova com presença total no ano de 2009.

O número de grupos definidos variava de acordo com a área de enquadramento e com a região que estava sendo observada, porém, evitou-se trabalhar com um número de grupos que resultavam em um volume de incidências menor que 3%. Em alguns casos isso não era possível, como o que ocorre na Figura 15, onde é possível ver que o grupo 4 possui um volume de incidência de apenas 1% do total.

4.1.6 Análise e validação dos Dados

Para a avaliação dos resultados, todas as saídas geradas pelo algoritmo na ferramenta Weka foram reunidos em planilhas do Microsoft Excel, a Figura 16 apresenta a forma como os dados ficaram organizados nas planilhas para uma das áreas de enquadramento. Foi feita também, através de fórmulas na planilha, a adição de colunas que correspondem à quantidade de questões total analisadas uma área e a quantidade de estudantes que acertaram, erraram ou deixaram em branco as questões. Tais dados também foram adicionados em porcentagem na planilha.

Figura 16 - Organização dos dados após mineração como algoritmo Simple KMeans

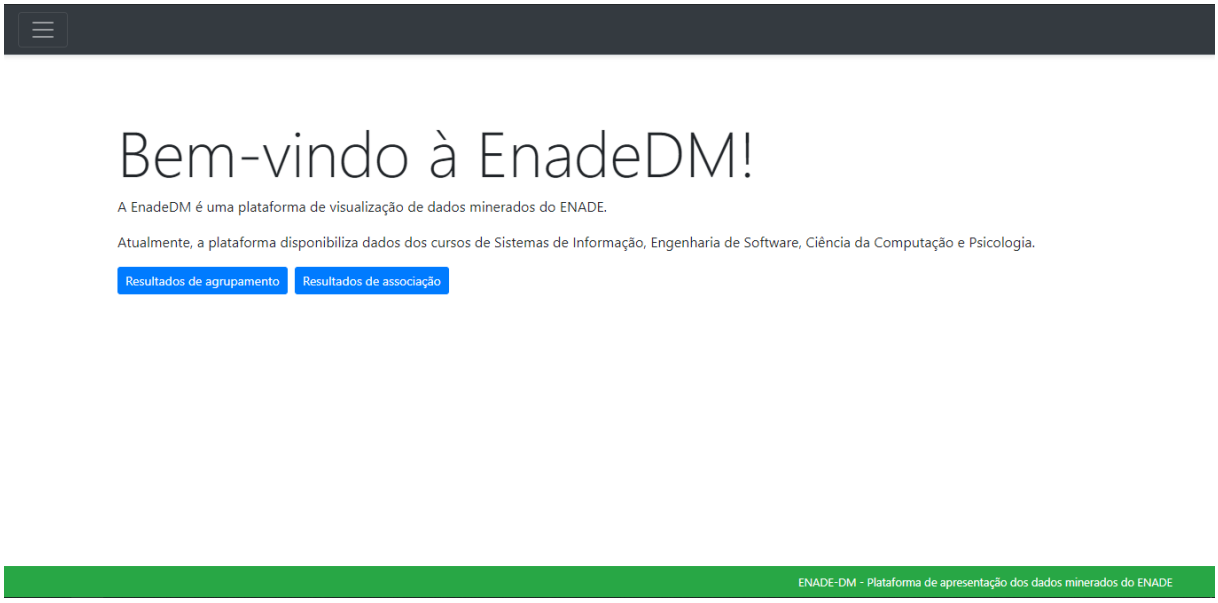
area_enquadramento	curso	ano	regiao	questao	questao	cluster	volume_incidencias	volume_incidencias_porcentagem	qtd_questoes	qtd_certas	qtd_erradas	qtd_branco_invalidas	porcentagem_certas	porcentagem_erradas	porcentagem_branco_invalido
História e Sistemas da Psicologia	Psicologia	2009	Norte	Certa	Certa	0	250	13%	2	2	0	0	100%	0%	0%
História e Sistemas da Psicologia	Psicologia	2009	Norte	Certa	Errada	1	726	38%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Norte	Errada	Errada	2	761	40%	2	0	2	0	0%	100%	0%
História e Sistemas da Psicologia	Psicologia	2009	Norte	X	Errada	3	17	1%	2	0	1	1	0%	50%	50%
História e Sistemas da Psicologia	Psicologia	2009	Norte	Errada	Certa	4	143	8%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Nordeste	Certa	Errada	0	2901	41%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Nordeste	Certa	Certa	1	1281	18%	2	2	0	0	100%	0%	0%
História e Sistemas da Psicologia	Psicologia	2009	Nordeste	Errada	Errada	2	2254	32%	2	0	2	0	0%	100%	0%
História e Sistemas da Psicologia	Psicologia	2009	Nordeste	Errada	Certa	3	568	8%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Nordeste	Errada	X	4	60	1%	2	0	1	1	0%	50%	50%
História e Sistemas da Psicologia	Psicologia	2009	Centro-Oeste	Certa	Errada	0	1413	41%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Centro-Oeste	Certa	Certa	1	662	19%	2	2	0	0	100%	0%	0%
História e Sistemas da Psicologia	Psicologia	2009	Centro-Oeste	Errada	Certa	2	296	9%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Centro-Oeste	Errada	Errada	3	1020	30%	2	0	2	0	0%	100%	0%
História e Sistemas da Psicologia	Psicologia	2009	Centro-Oeste	X	X	4	14	0%	2	0	0	2	0%	0%	100%
História e Sistemas da Psicologia	Psicologia	2009	Sudeste	Certa	Errada	0	8475	42%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sudeste	Errada	Errada	1	6990	35%	2	0	2	0	0%	100%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sudeste	Certa	Certa	2	2862	14%	2	2	0	0	100%	0%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sudeste	Errada	Certa	3	1696	8%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sudeste	X	Errada	4	142	1%	2	0	1	1	0%	50%	50%
História e Sistemas da Psicologia	Psicologia	2009	Sul	Certa	Errada	0	3479	44%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sul	Errada	Errada	1	2547	32%	2	0	2	0	0%	100%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sul	Errada	Certa	2	662	8%	2	1	1	0	50%	50%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sul	Certa	Certa	3	1097	14%	2	2	0	0	100%	0%	0%
História e Sistemas da Psicologia	Psicologia	2009	Sul	Certa	X	4	99	1%	2	1	0	1	50%	0%	50%

Através dos dados apresentados na figura acima, já é possível verificar algumas informações e realizar comparativos. Para a o conteúdo de História e Sistemas da Psicologia em 2009, na região Norte, há um grupo de 250 estudantes que acertaram as duas questões correspondentes ao conteúdo curricular, o que corresponde a 13% do total da região. Já na região Nordeste, a quantidade de alunos que acertaram as duas questões é de 1281, 18% do total de estudantes que realizaram a prova com presença total na região.

4.1.7 Plataforma de Visualização

Os dados tratados e minerados nos passos anteriores foram incluídos na plataforma de visualização de dados minerados do ENADE, que já contava com a disponibilidade de resultados da área de Computação. A Figura 17 mostra a tela inicial da plataforma, que disponibiliza duas opções: a visualização dos dados minerados com a tarefa de agrupamento ou com a de associação.

Figura 17 - Tela inicial da plataforma EnadeDM



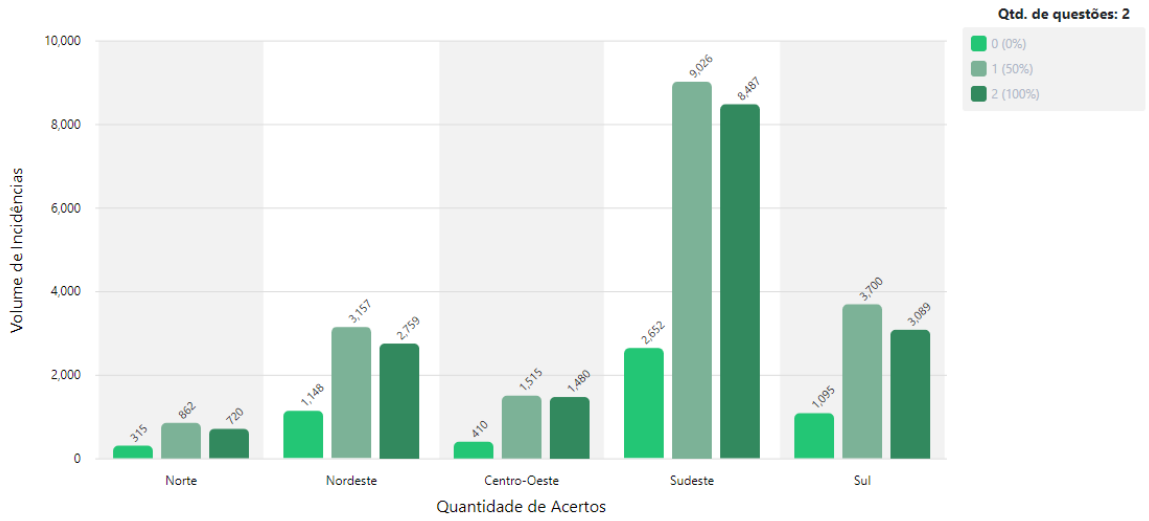
Ao clicar para acessar os resultados de agrupamento é possível ver os filtros possíveis para os dados disponíveis. A Figura 18 apresenta a tela inicial dos gráficos antes de fazer uma busca. Nos filtros estão disponíveis os cursos de Sistemas de Informação, Engenharia de Software, Ciência da Computação e Psicologia, além de seus respectivos anos e área de enquadramento de questões.

Figura 18 - Tela de filtros da plataforma

Após realizar a busca por um dos filtros, a plataforma exibirá os resultados em formato de gráficos. O gráfico da Figura 19 apresenta a quantidade de acertos para as duas questões de uma das áreas de enquadramento. Do lado direito, a legenda informa quais barras correspondem

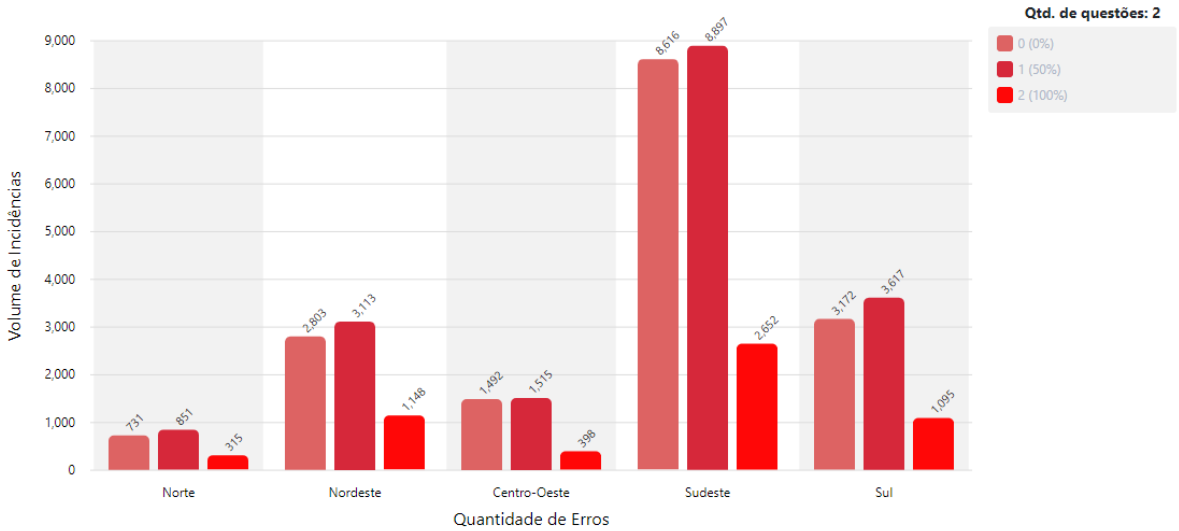
à quantidade de alunos que acertaram zero - das duas questões - os que acertaram uma e os que acertaram as duas, é possível ver o volume de incidências através dos gráficos.

Figura 19 – Gráfico de acertos em barra



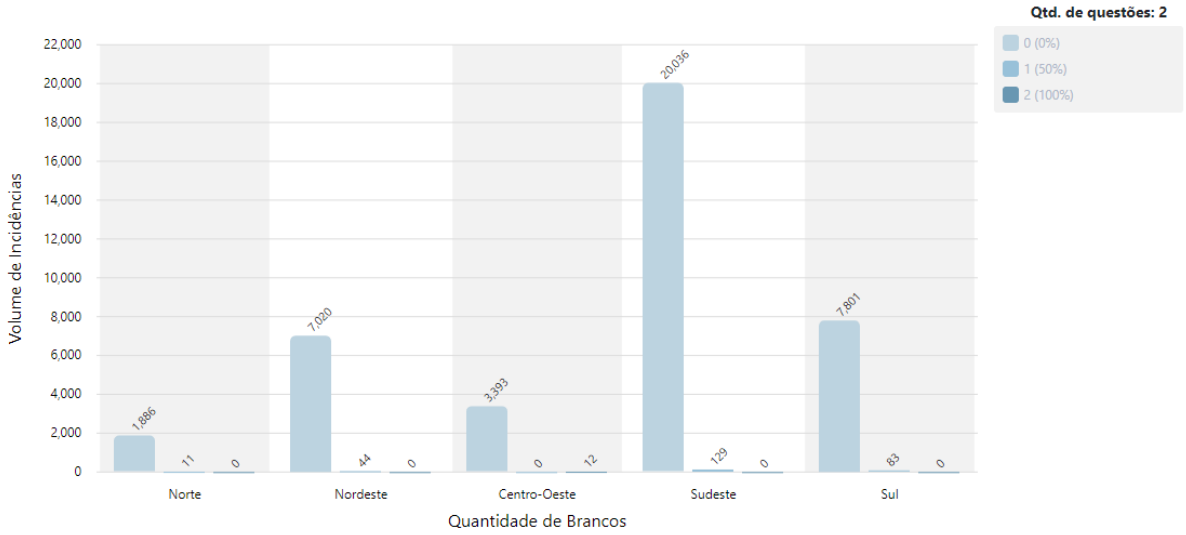
O gráfico da Figura 20 tem o mesmo sentido do gráfico acima, porém, nele são exibidas as quantidades de erros, os que erraram nenhuma, uma e erraram as duas questões da área de enquadramento.

Figura 20 – Gráfico de erros em barra



No gráfico da Figura 21 são exibidas as quantidades de estudantes que deixaram as questões em branco ou que tiveram as mesmas anuladas.

Figura 21 - Gráfico de brancos/nulos em barra



É possível ver que, para todas as regiões, exceto o Centro-Oeste, não houve estudantes que deixaram as duas questões em branco ou tiveram as duas anuladas, nota-se que, em grande parte, praticamente todos os alunos de todas as regiões responderam as duas questões e não tiveram suas respostas anuladas.

4.2 RESULTADOS DE ASSOCIAÇÃO

Esta seção apresenta os resultados obtidos através da utilização da tarefa de associação. Nesta etapa foram selecionados somente estudantes que fizeram a prova do ENADE por instituições de ensino que obtiveram CPC 4 e 5, buscando traçar os perfis desses estudantes que levaram a instituição a alcançar as maiores notas.

4.2.1 Extração dos Dados

Considerando que nesta etapa o objetivo era analisar os perfis dos estudantes das instituições de ensino com Conceito Preliminar de Curso (CPC), fez-se necessário, além da coleta dos dados dos estudantes que participaram do ENADE, a busca dos dados com informações de cada instituição.

Figura 22 - Tabela com informações de IES participantes do ENADE 2015

Ano	Código da IES	Nome da IES	Sigla da IES	Organização Acadêmica	Categoria Administrativa	Código do Curso	Código da Área	Área de Enquadramento	Modalidade de Ensino	Código do Município	Município do Curso	Sigla da UF	Concluintes Inscritos	Concluintes Participantes	Nota Bruta - FG	Nota Bruta - CE
2015	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	UFMT	Universidade	Federal	13	1	ADMINISTRAÇÃO	Educação Presencial	5103403	CUIABA	MT	76	66	66,6909	51,4167
2015	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	UFMT	Universidade	Federal	1127343	1	ADMINISTRAÇÃO	Educação Presencial	5107602	RONDONOPOLIS	MT	22	22	58,1682	42,2955
2015	2	UNIVERSIDADE DE BRASÍLIA	UNB	Universidade	Federal	145	1	ADMINISTRAÇÃO	Educação Presencial	5300108	BRASILIA	DF	232	157	65,9943	56,0401
2015	3	UNIVERSIDADE FEDERAL DE SERGIPE	UFS	Universidade	Federal	305	1	ADMINISTRAÇÃO	Educação Presencial	2806701	SAO CRISTOVAO	SE	109	86	52,8500	40,0081
2015	3	UNIVERSIDADE FEDERAL DE SERGIPE	UFS	Universidade	Federal	20762	1	ADMINISTRAÇÃO	Educação Presencial	2806701	SAO CRISTOVAO	SE	86	54	54,3315	40,9130
2015	3	UNIVERSIDADE FEDERAL DE SERGIPE	UFS	Universidade	Federal	95003	1	ADMINISTRAÇÃO	Educação Presencial	2802908	ITABAIANA	SE	54	37	55,4351	37,8324
2015	4	UNIVERSIDADE FEDERAL DO AMAZONAS	UFAM	Universidade	Federal	392	1	ADMINISTRAÇÃO	Educação Presencial	1302603	MANAUS	AM	183	154	61,3708	45,1617
2015	4	UNIVERSIDADE FEDERAL DO AMAZONAS	UFAM	Universidade	Federal	101608	1	ADMINISTRAÇÃO	Educação Presencial	1300607	BENJAMIN CONSTA	AM	66	60	53,2217	37,4800
2015	4	UNIVERSIDADE FEDERAL DO AMAZONAS	UFAM	Universidade	Federal	112106	1	ADMINISTRAÇÃO	Educação Presencial	1303403	PARINTINS	AM	75	60	56,7467	35,5167
2015	4	UNIVERSIDADE FEDERAL DO AMAZONAS	UFAM	Universidade	Federal	1104306	1	ADMINISTRAÇÃO	Educação a Distância	1302603	MANAUS	AM	17	1		
2015	5	UNIVERSIDADE FEDERAL DO PIAUÍ	UFPI	Universidade	Federal	500	1	ADMINISTRAÇÃO	Educação Presencial	2207702	PARNAIBA	PI	36	28	53,5536	37,4357
2015	5	UNIVERSIDADE FEDERAL DO PIAUÍ	UFPI	Universidade	Federal	18373	1	ADMINISTRAÇÃO	Educação Presencial	2211001	TERESINA	PI	87	57	62,6912	45,9702
2015	5	UNIVERSIDADE FEDERAL DO PIAUÍ	UFPI	Universidade	Federal	99870	1	ADMINISTRAÇÃO	Educação Presencial	2208007	PICOS	PI	46	35	54,4857	38,0143
2015	5	UNIVERSIDADE FEDERAL DO PIAUÍ	UFPI	Universidade	Federal	108006	1	ADMINISTRAÇÃO	Educação a Distância	2211001	TERESINA	PI	9	0		
2015	5	UNIVERSIDADE FEDERAL DO PIAUÍ	UFPI	Universidade	Federal	116640	1	ADMINISTRAÇÃO	Educação Presencial	2203909	FLORIANO	PI	109	69	54,5174	37,0130
2015	6	UNIVERSIDADE FEDERAL DE OURO PRETO	UFOP	Universidade	Federal	112594	1	ADMINISTRAÇÃO	Educação Presencial	3140001	MARIANA	MG	98	85	60,5800	44,3388
2015	7	UNIVERSIDADE FEDERAL DE SÃO CARLOS	UFSCAR	Universidade	Federal	115084	1	ADMINISTRAÇÃO	Educação Presencial	3552205	SOROCABA	SP	69	64	66,6891	59,8531
2015	8	UNIVERSIDADE FEDERAL DE VIÇOSA	UFV	Universidade	Federal	683	1	ADMINISTRAÇÃO	Educação Presencial	3171303	VICOSA	MG	79	74	66,5095	55,8932
2015	8	UNIVERSIDADE FEDERAL DE VIÇOSA	UFV	Universidade	Federal	111120	1	ADMINISTRAÇÃO	Educação Presencial	3155504	RIO PARANAIBA	MG	42	39	59,7385	45,0769
2015	8	UNIVERSIDADE FEDERAL DE VIÇOSA	UFV	Universidade	Federal	1126514	1	ADMINISTRAÇÃO	Educação Presencial	3126000	FLORESTAL	MG	42	39	54,7000	42,6077
2015	9	UNIVERSIDADE ESTADUAL DE LONDRINA	UEL	Universidade	Estadual	768	1	ADMINISTRAÇÃO	Educação Presencial	4113700	LONDRINA	PR	157	141	51,1340	51,3121
2015	10	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	PUCPR	Universidade	Privada	877	1	ADMINISTRAÇÃO	Educação Presencial	4125506	SAO JOSE DO PINHA	PR	35	29	59,1448	54,7310
2015	10	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	PUCPR	Universidade	Privada	19312	1	ADMINISTRAÇÃO	Educação Presencial	4115200	MARINGA	PR	25	24	55,4167	48,6375
2015	10	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	PUCPR	Universidade	Privada	21596	1	ADMINISTRAÇÃO	Educação Presencial	4106902	CURITIBA	PR	214	183	54,0852	45,8978
2015	10	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	PUCPR	Universidade	Privada	43543	1	ADMINISTRAÇÃO	Educação Presencial	4113700	LONDRINA	PR	49	43	58,4209	47,1047
2015	10	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	PUCPR	Universidade	Privada	60114	1	ADMINISTRAÇÃO	Educação Presencial	4127700	TOLEDO	PR	31	28	55,2036	42,5714
2015	11	UNIVERSIDADE CATÓLICA DE PERNAMBUCO	UNICAP	Universidade	Privada	960	1	ADMINISTRAÇÃO	Educação Presencial	2611606	RECIFE	PE	212	188	54,5968	39,3814
2015	11	UNIVERSIDADE CATÓLICA DE PERNAMBUCO	UNICAP	Universidade	Privada	29332	1	ADMINISTRAÇÃO	Educação Presencial	2611606	RECIFE	PE	9	4	74,0500	35,4500
2015	12	UNIVERSIDADE FEDERAL DO RIO GRANDE	FURG	Universidade	Federal	1028	1	ADMINISTRAÇÃO	Educação Presencial	4315602	RIO GRANDE	RS	28	24	64,0375	54,1792
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	31041	1	ADMINISTRAÇÃO	Educação Presencial	4305108	CAXIAS DO SUL	RS	520	468	51,6902	37,6293
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	32252	1	ADMINISTRAÇÃO	Educação Presencial	4322509	VACARIA	RS	48	39	52,4051	38,2564
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	34655	1	ADMINISTRAÇÃO	Educação Presencial	4307906	FARROUPILHA	RS	72	63	55,1698	41,0937
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	40435	1	ADMINISTRAÇÃO	Educação Presencial	4304408	CANELA	RS	76	57	52,5877	40,6140
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	40604	1	ADMINISTRAÇÃO	Educação Presencial	4313300	NOVA PRATA	RS	73	68	51,0441	34,5324
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	40943	1	ADMINISTRAÇÃO	Educação Presencial	4302105	BENTO GONCALVES	RS	115	90	52,3422	37,7389
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	42313	1	ADMINISTRAÇÃO	Educação Presencial	4319505	SAO SEBASTIAO DO	RS	103	91	54,9725	36,8791
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	68237	1	ADMINISTRAÇÃO	Educação Presencial	4309407	GUAPORE	RS	48	45	47,4044	32,9578
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	106910	1	ADMINISTRAÇÃO	Educação Presencial	4305108	CAXIAS DO SUL	RS	106	92	54,3413	37,6272
2015	13	UNIVERSIDADE DE CAXIAS DO SUL	UCS	Universidade	Privada	106912	1	ADMINISTRAÇÃO	Educação Presencial	4302105	BENTO GONCALVES	RS	53	48	58,1042	39,1438

A Figura 22 apresenta as instituições de ensino superior que foram avaliadas no ENADE. A planilha contém informações gerais de cada IES, como estado, município e o CPC.

4.2.2 Entendimento dos Dados

Levando em conta que o foco do trabalho foi no curso de Psicologia, foram excluídas da planilha as IES que não tinham curso nesta área. Também foram removidas as IES que não apresentavam valor 4 ou 5 na coluna `cpc_faixa`.

Figura 23 - IES com CPC 4 e 5 do curso de Psicologia

ano_enade	cod_ies	ies	sigla	organizacao	tipo	cod_area	area	cod_municipio	municipio	estado	cpc_faixa
2015	1	UNIVERSIDADE FEDERAL DE MATO GROSSO	UFMT	Universidade	Federal	18	PSICOLOGIA	5107602	RONDONOPOLIS	MT	4
2015	2	UNIVERSIDADE DE BRASÍLIA	UNB	Universidade	Federal	18	PSICOLOGIA	5300108	BRASILIA	DF	4
2015	2	UNIVERSIDADE DE BRASÍLIA	UNB	Universidade	Federal	18	PSICOLOGIA	5300108	BRASILIA	DF	4
2015	3	UNIVERSIDADE FEDERAL DE SERGIPE	UFS	Universidade	Federal	18	PSICOLOGIA	2806701	SAO CRISTOVAO	SE	4
2015	4	UNIVERSIDADE FEDERAL DO AMAZONAS	UFAM	Universidade	Federal	18	PSICOLOGIA	1302603	MANAUS	AM	4
2015	7	UNIVERSIDADE FEDERAL DE SÃO CARLOS	UFSCAR	Universidade	Federal	18	PSICOLOGIA	3548906	SAO CARLOS	SP	4
2015	9	UNIVERSIDADE ESTADUAL DE LONDRINA	UEL	Universidade	Estadual	18	PSICOLOGIA	4113700	LONDRINA	PR	4
2015	9	UNIVERSIDADE ESTADUAL DE LONDRINA	UEL	Universidade	Estadual	18	PSICOLOGIA	4113700	LONDRINA	PR	4
2015	11	UNIVERSIDADE CATÓLICA DE PERNAMBUCO	UNICAP	Universidade	Privada	18	PSICOLOGIA	2611606	RECIFE	PE	4
2015	14	UNIVERSIDADE DO VALE DO RIO DOS SINOS	UNISINOS	Universidade	Privada	18	PSICOLOGIA	4318705	SAO LEOPOLDO	RS	4
2015	17	UNIVERSIDADE FEDERAL DE UBERLÂNDIA	UFU	Universidade	Federal	18	PSICOLOGIA	3170206	UBERLANDIA	MG	4
2015	18	UNIVERSIDADE CATÓLICA DE PELOTAS	UCPEL	Universidade	Privada	18	PSICOLOGIA	4314407	PELOTAS	RS	4
2015	22	UNIVERSIDADE PRESBITERIANA MACKENZIE	MACKENZIE	Universidade	Privada	18	PSICOLOGIA	3550308	SAO PAULO	SP	4
2015	23	UNIVERSIDADE FEEVALE	FEEVALE	Universidade	Privada	18	PSICOLOGIA	4313409	NOVO HAMBURGO	RS	4
2015	27	UNIVERSIDADE VALE DO RIO VERDE	UNINCOR	Universidade	Privada	18	PSICOLOGIA	3163307	TRES CORACOES	MG	4
2015	57	UNIVERSIDADE ESTADUAL DE MARINGÁ	UEM	Universidade	Estadual	18	PSICOLOGIA	4115200	MARINGA	PR	4
2015	60	CENTRO UNIVERSITÁRIO MUNICIPAL DE FRANCA	UNI-FACEF	Centro Universitário	Estadual	18	PSICOLOGIA	3516200	FRANCA	SP	4
2015	82	UNIVERSIDADE DO OESTE DE SANTA CATARINA	UNOESC	Universidade	Estadual	18	PSICOLOGIA	4217204	SAO MIGUEL DOESTE	SC	4
2015	107	UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL REI	UFSJ	Universidade	Federal	18	PSICOLOGIA	3162500	SAO JOAO DEL REI	MG	4
2015	138	CENTRO UNIVERSITÁRIO BARÃO DE MAUÁ	CBM	Centro Universitário	Privada	18	PSICOLOGIA	3543402	RIBEIRAO PRETO	SP	4
2015	163	UNIVERSIDADE ESTÁCIO DE SÁ	UNESA	Universidade	Privada	18	PSICOLOGIA	3303401	NOVA FRIBURGO	RJ	4
2015	163	UNIVERSIDADE ESTÁCIO DE SÁ	UNESA	Universidade	Privada	18	PSICOLOGIA	3301009	CAMPOS DO GOYTACAZES	RJ	4
2015	163	UNIVERSIDADE ESTÁCIO DE SÁ	UNESA	Universidade	Privada	18	PSICOLOGIA	3303302	NITEROI	RJ	4
2015	163	UNIVERSIDADE ESTÁCIO DE SÁ	UNESA	Universidade	Privada	18	PSICOLOGIA	3304557	RIO DE JANEIRO	RJ	4
2015	163	UNIVERSIDADE ESTÁCIO DE SÁ	UNESA	Universidade	Privada	18	PSICOLOGIA	3304557	RIO DE JANEIRO	RJ	4
2015	163	UNIVERSIDADE ESTÁCIO DE SÁ	UNESA	Universidade	Privada	18	PSICOLOGIA	3303500	NOVA IGUACU	RJ	4
2015	165	UNIVERSIDADE VEIGA DE ALMEIDA	UVA	Universidade	Privada	18	PSICOLOGIA	3304557	RIO DE JANEIRO	RJ	4
2015	165	UNIVERSIDADE VEIGA DE ALMEIDA	UVA	Universidade	Privada	18	PSICOLOGIA	3300704	CABO FRIO	RJ	4

A Figura 23 mostra a planilha após as alterações realizadas. De forma geral, foram mantidas as colunas

- ANO_ENADE: coluna que contém o ano correspondente da avaliação da IES pelo ENADE;
- COD_IES: código da IES definido pela organização do ENADE;
- IES: Nome da instituição
- SIGLA: sigla que representa o nome completo da IES;
- ORGANIZAÇÃO: apresenta o tipo de organização da IES, exemplo: Universidade ou Faculdade;

- TIPO: o nível da organização da IES, exemplo: Federal, Estadual ou Privada;
- COD_AREA: código que representa a área, foram mantidos somente as linhas com código 18, que corresponde a área de Psicologia;
- AREA: a área avaliada na IES, neste caso, a área de Psicologia;
- COD_MUNICIPIO: o código que representa o munícipio no qual a IES está localizada;
- MUNICIPIO: o nome do município onde a IES se encontra;
- ESTADO: o estado correspondente à localização da IES;
- CPC_FAIXA: essa coluna apresenta o Conceito Preliminar do Curso que determinada instituição obteve para cada curso.

Na planilha com dados dos estudantes, para o ano de 2009 e 2012, foram mantidas colunas com algumas informações pessoais, da IES na qual está matriculado e resposta de algumas questões do questionário de perguntas socioeconômicas:

- ANO: o ano em que o aluno participou da prova do ENADE;
- COD_IES: código da instituição do estudante;
- TIPO: o nível da organização da IES, exemplo: Federal, Estadual ou Privada;
- ORGANIZAÇÃO: apresenta o tipo de organização da IES, exemplo: Universidade ou Faculdade;
- CURSO: o curso do estudante;
- COD_MUNICIPIO: código do município localizada a IES do estudante;
- COD_ESTADO: o código do estado onde está localizada a IES;
- REGIAO: a região do país onde se encontra a IES;
- SEXO: sexo do estudante que participou do exame;

Também foram mantidas algumas respostas dos estudantes do questionário socioeconômico. Estas serviram para agregar os perfis dos estudantes durante o estudo.

Figura 24 - Informações de estudantes conservadas nos anos de 2009 e 2012

ano	cod_ies	tipo	organizacao	curso	regiao	cod_estado	cod_municipio	sexo	estado	QE_I9	QE_I17	QE_I35	QE_I46	QE_I50
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=B	QE_17=B	QE_35=A	QE_46=D	QE_50=B
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=B	QE_17=A	QE_35=A	QE_46=D	QE_50=B
2009	1	1	1	Psicologia	5	51	5107602	1	MT	QE_9=B	QE_17=C	QE_35=A	QE_46=A	QE_50=C
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=B	QE_17=B	QE_35=C	QE_46=D	QE_50=A
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=B	QE_17=C	QE_35=A	QE_46=D	QE_50=B
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=B	QE_17=B	QE_35=B	QE_46=A	QE_50=B
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=B	QE_17=B	QE_35=B	QE_46=D	QE_50=B
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=C	QE_17=B	QE_35=B	QE_46=A	QE_50=B
2009	1	1	1	Psicologia	5	51	5107602	2	MT	QE_9=C	QE_17=D	QE_35=B	QE_46=A	QE_50=B

A Figura 24 apresenta as informações de cada aluno mantidas para trabalhar com os anos de 2009 e 2012. Para o ano de 2015 foram mantidas as informações, porém com diferenças de questões do questionário socioeconômico. As questões e as alternativas correspondentes estão disponíveis nos apêndices do trabalho.

As colunas mostradas são as informações dos alunos que participaram da prova do ENADE, estas serviram para realizar a associação ele os dados de um estudante e o nível de CPC 4 ou 5.

4.2.3 Adequações da Proposta

Inicialmente, havia sido definido o foco do trabalho somente em cima das repostas das provas realizadas pelos estudantes, com intuito de utilizar das tarefas de agrupamento e classificação como tentativa de obtenção de resultados. Foram alcançados resultados através da tarefa de agrupamento, como apresentado na seção 4.1, porém, os dados não permitiram a prática com a tarefa de classificação, uma vez que esta exige um conjunto de treinamento para predição de novos valores, e os dados disponíveis não se adaptavam a tal cenário.

Portanto, foi definido junto à orientadora do projeto, a busca de novos conhecimentos com a tarefa de associação, com objetivo de verificar os perfis dos estudantes das IES com CPC 4 e 5. Para conhecimento da tarefa de associação, fez-se necessária a inclusão de referências teóricas na seção 2.

4.2.4 Adequações dos Dados

Antes da utilização dos algoritmos em busca de resultados, foi necessário adicionar algumas informações na planilha com os dados de cada aluno. A

Figura **25** apresenta a planilha com informações dos alunos, foram inseridas novas colunas que foram preenchidas pelo cruzamento com os dados da planilha com informações das IES contidas na planilha exibida na Figura 23.

Figura 25 - Dados de cada aluno antes do cruzamento com a lista de IES

ano	cod_ies	sigla	ies	tipo	organizacao	curso	cod_municipio	municipio	cod_estado	estado	regiao	sexo	QE_10	QE_11	QE_14	QE_15	QE_18	conceito_enade
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	M	QE_10=A	QE_11=A	QE_14=A	QE_15=B	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	M	QE_10=B	QE_11=A	QE_14=A	QE_15=D	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=C	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=D	QE_11=A	QE_14=A	QE_15=F	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=B	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9			115	10028	Psicologia	4113700		41		4	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	

O código mostrado na Figura 26 abaixo passa por cada linha da planilha de estudantes exibida acima, troca o código da coluna `regiao` pelo nome da região correspondente. Após isso, entra na planilha com dados das IES e procura a que possui a coluna `cod_ies` com o mesmo valor da coluna `cod_ies` da planilha de alunos, também faz o mesmo para a coluna de código do município. Após encontrar a condição citada, o código move as informações contidas nas colunas `conceito_enade`, `municipio`, `organizacao`, `tipo`, `ies`, `sigla` e `estado` para as colunas de mesmo nome presentes na planilha de estudantes.

Figura 26 - *Script* para cruzamento das informações das IES com estudantes

```
alunos = pd.read_excel('psico_2015_separados.xls')
ies = pd.read_excel('conceito_2015.xls')

tamanho_alunos = len(alunos)
tamanho_ies = len(ies)

for i in range(tamanho_alunos):
    if alunos['regiao'][i] == 1:
        alunos['regiao'][i] = 'Norte'
    elif alunos['regiao'][i] == 2:
        alunos['regiao'][i] = 'Nordeste'
    elif alunos['regiao'][i] == 3:
        alunos['regiao'][i] = 'Sudeste'
    elif alunos['regiao'][i] == 4:
        alunos['regiao'][i] = 'Sul'
    elif alunos['regiao'][i] == 5:
        alunos['regiao'][i] = 'Centro-Oeste'
    for j in range(tamanho_ies):
        if alunos['cod_ies'][i] == ies['cod_ies'][j] and alunos['cod_municipio'][i] == ies['cod_municipio'][j]:
            alunos['conceito_enade'][i] = ies['conceito_enade'][j]
            alunos['municipio'][i] = ies['municipio'][j]
            alunos['organizacao'][i] = ies['organizacao'][j]
            alunos['tipo'][i] = ies['tipo'][j]
            alunos['ies'][i] = ies['ies'][j]
            alunos['sigla'][i] = ies['sigla'][j]
            alunos['estado'][i] = ies['estado'][j]
```

Após a execução do código acima, é gerada uma nova planilha com informações completas dos estudantes de IES com CPC 4 e 5. Esta nova planilha está exibida na Figura 27.

Figura 27 - Planilha com dados preenchidos de alunos de IES com CPC 4 e 5

ano	cod_ies	sigla	ies	tipo	organizacao	curso	cod_municipio	municipio	cod_estado	estado	regiao	sexo	QE_10	QE_11	QE_14	QE_15	QE_18	conceito_enade
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=A	QE_11=A	QE_14=A	QE_15=B	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=B	QE_11=A	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=C	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=D	QE_11=A	QE_14=A	QE_15=F	QE_18=A	4
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=B	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	M	QE_10=B	QE_11=A	QE_14=A	QE_15=A	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	M	QE_10=A	QE_11=A	QE_14=A	QE_15=E	QE_18=A	
2015	9				Universidade	Psicologia	4113700		41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	

Nota-se que restaram linhas que não possuem todas as informações. Tais linhas representam alunos que não fizeram parte de IES que alcançaram CPC 4 ou 5, portanto, os

mesmos foram removidos da planilha, mantendo apenas os estudantes das IES com CPC mais alto.

Após a realização dos passos anteriores, foram removidas as colunas que apresentavam apenas códigos, pois estas colunas não agregariam para a utilização da mineração de dados. A Figura 28 apresenta os dados com as colunas removidas.

Figura 28 - Dados de associação prontos para mineração

ano	cod_ies	sigla	ies	tipo	organizacao	curso	cod_municipio	municipio	cod_estado	estado	regiao	sexo	QE_10	QE_11	QE_14	QE_15	QE_18	conceito_enade
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=A	QE_11=A	QE_14=A	QE_15=B	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=B	QE_11=A	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=C	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=D	QE_11=A	QE_14=A	QE_15=F	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=B	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=B	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=A	QE_11=A	QE_14=A	QE_15=E	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=H	QE_14=A	QE_15=D	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=C	QE_11=H	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	F	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=B	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4
2015	9	UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	Estadual	Universidade	Psicologia	4113700	LONDRINA	41	PR	Sul	M	QE_10=A	QE_11=A	QE_14=A	QE_15=A	QE_18=A	4

Com todas as adequações de dados realizadas, os dados estavam prontos para serem minerados com a utilização de recursos da tarefa de associação. A demonstração da mineração dos dados está contida na seção seguinte.

4.2.5 Aplicação dos Algoritmos

Para a mineração dos dados nesta etapa do trabalho foi utilizado o conceito de associação, uma das principais tarefas de mineração de dados, juntamente com o algoritmo de Apriori na linguagem de programação R. Para a utilização do algoritmo com a linguagem R, é necessário instalar e importar para o ambiente do R Studio o pacote “arules”.

Figura 29 - Instalação e importação do pacote "arules"

```
install.packages('arules')
library(arules)
```

A Figura 29 apresenta duas linhas de código da linguagem R. Na primeira, o programa executa a instalação do pacote necessário para utilização do algoritmo de Apriori e, na segunda linha, importa o mesmo pacote para o ambiente do R Studio.

Figura 30 - Importando os dados de estudantes para uso no Apriori

```

txn = read.transactions(file="psico_2015_minerar.csv", format="basket",sep=";", cols = NULL);

```

Após a instalação e importação do pacote para utilização do algoritmo Apriori, foi necessário carregar a planilha com as informações dos estudantes para mineração. Na Figura 30 está o código necessário para tal. O que o trecho de código faz é, simplesmente, ler o arquivo de formato .csv e atribuir em uma variável denominada “txn”.

Figura 31 - Buscando regras de associação nos dados com informações de estudantes

```
rules <- apriori(txn, parameter = list(sup = 0.4, conf = 0.9,target="rules"),
                 appearance = list(rhs=c("4", "5"),
                                   default="lhs"));
```

O código da Figura 31 é o responsável por buscar as regras de associação. Através da utilização da função “apriori()”, é possível passar a lista de dados que deseja-se minerar, os parâmetros de suporte e confiança e também o consequente que pretende-se alcançar, que, neste estudo, foram os CPC 4 e 5. O resultados encontrados pela execução da função são atribuídos para a variável “rules”, e podem ser visualizados através da execução da função “inspect(rules)”.

Figura 32 - Regras de associação geradas pelo algoritmo de Apriori

```
{QE_14=A,Sudeste,Universidade}        => {4} 0.4006374 1.0000000
{QE_14=A,QE_15=A,Sudeste}             => {4} 0.4050581 1.0000000
{F,QE_14=A,Sudeste}                   => {4} 0.4121517 1.0000000
{QE_14=A,QE_18=A,Sudeste}             => {4} 0.4143107 1.0000000
{F,QE_10=A,QE_15=A}                   => {4} 0.4039272 0.9961968
{QE_10=A,QE_15=A,QE_18=A}             => {4} 0.4023851 0.9966896
{QE_10=A,QE_14=A,QE_15=A}             => {4} 0.4405264 0.9965116
{F,QE_10=A,QE_18=A}                   => {4} 0.4309654 0.9969084
```

A Figura 32 exibe algumas regras geradas pelo algoritmo Apriori com os perfis de estudantes de Psicologia do ano de 2015. As informações podem ser mais bem compreendidas com o acompanhamento das informações das questões do questionário do estudante apresentados na seção 4.2.2.

4.2.6 Análise e validação dos Dados

Conforme visto na Figura 32, os resultados de associação gerados pelo Apriori permitem tomar conhecimento quanto aos perfis dos estudantes de IES com CPC 4 no ENADE. Por exemplo, na primeira regra: estudantes que afirmaram não ter participado de atividades ou programas no exterior durante a graduação – informação que corresponde à alternativa A da questão 14 do questionário do estudante - e são de uma IES universitária da região sudeste do Brasil, estão em uma instituição com CPC 4. Tal padrão ocorre em 40% das vezes.

Nota-se, em outras regras, a ocorrência – em torno de 40% - de estudantes que afirmam não ter participado de programas fora do país durante a graduação e também estão em

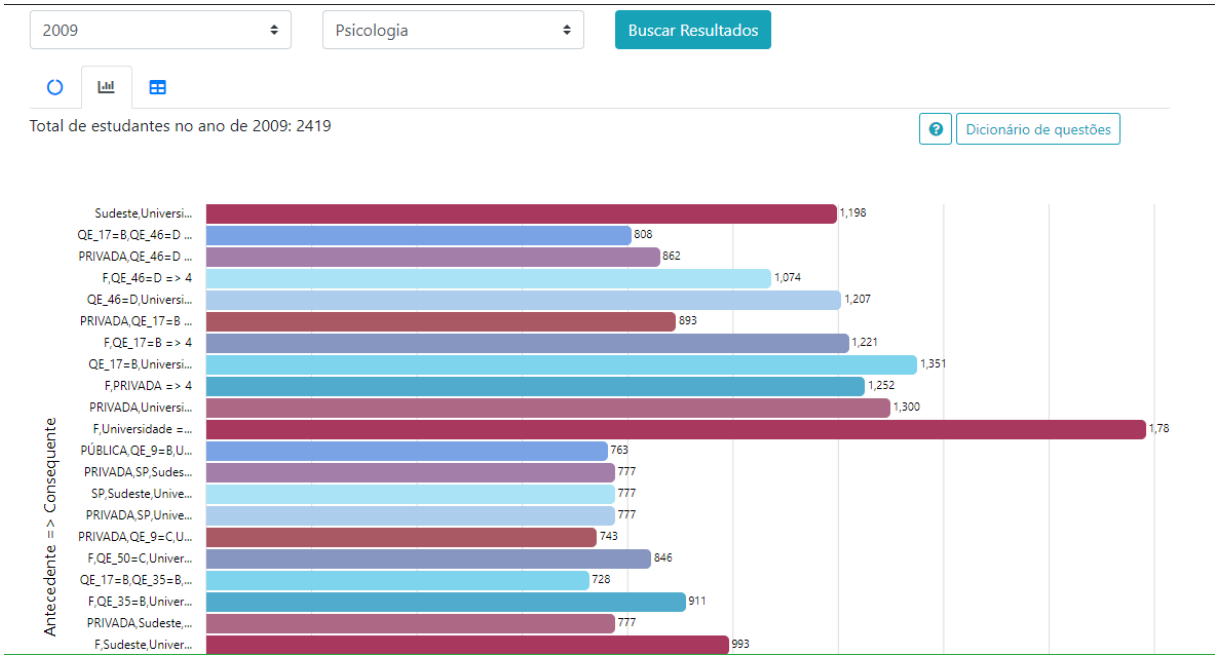
instituições com CPC 4, o que pode significar que o fato desses estudantes não terem realizado atividades no exterior não faz com que determinada IES tenha baixo desempenho nas provas do ENADE.

Durante a mineração dos dados, o algoritmo não foi capaz de gerar regras com consequente equivalente ao CPC 5, pelo fato de o número de IES que tiraram nota máxima ser menor, então o Apriori não conseguiu encontrar padrões nos perfis dos alunos de IES com CPC máximo. Foi feita a tentativa de minerar somente os dados de estudantes das IES com conceito 5, porém, ainda assim, não foi possível encontrar padrões de associação nos dados.

4.2.7 Plataforma de Visualização

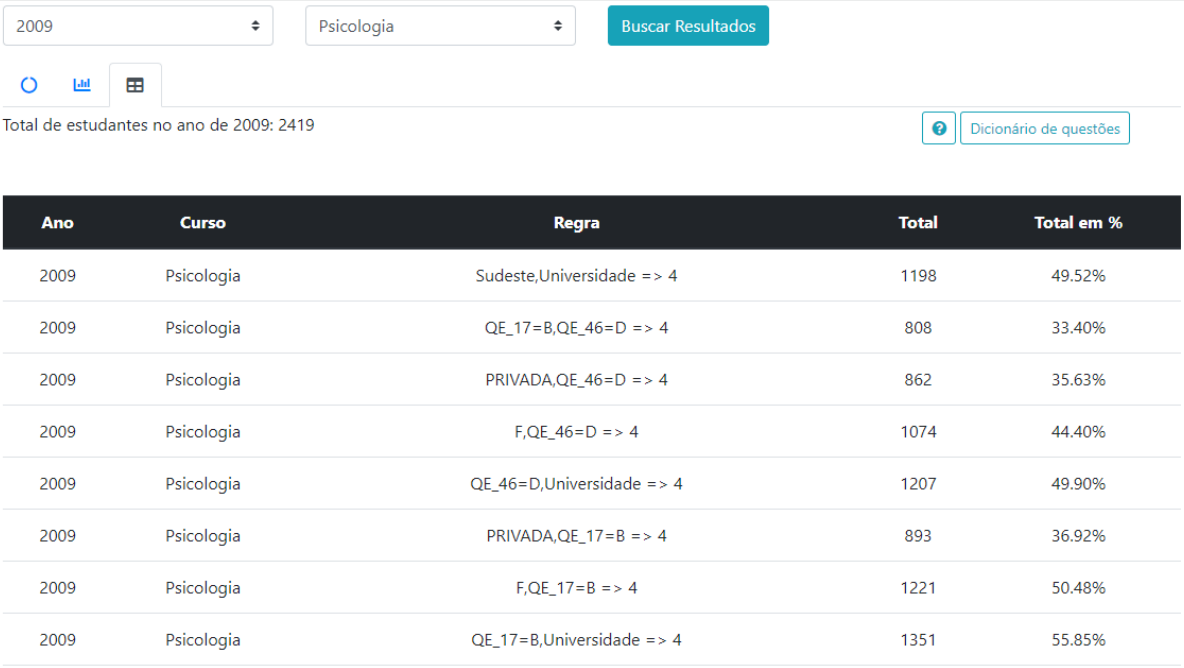
Para a visualização dos resultados obtidos através da mineração de dados com a tarefa de associação foi criado um módulo na plataforma EnadeDM. Ao acessar o *link* de "Resultados de associação" na tela inicial ou no menu superior, o sistema irá exibir os resultados desejados.

Figura 33 - Gráfico de visualização dos resultados de associação



A Figura 33 apresenta parte de um gráfico com as regras de associação geradas pelo algoritmo Apriori. Para cada regra é possível ver também a quantidade de estudantes cujo perfil se encaixa naquela regra.

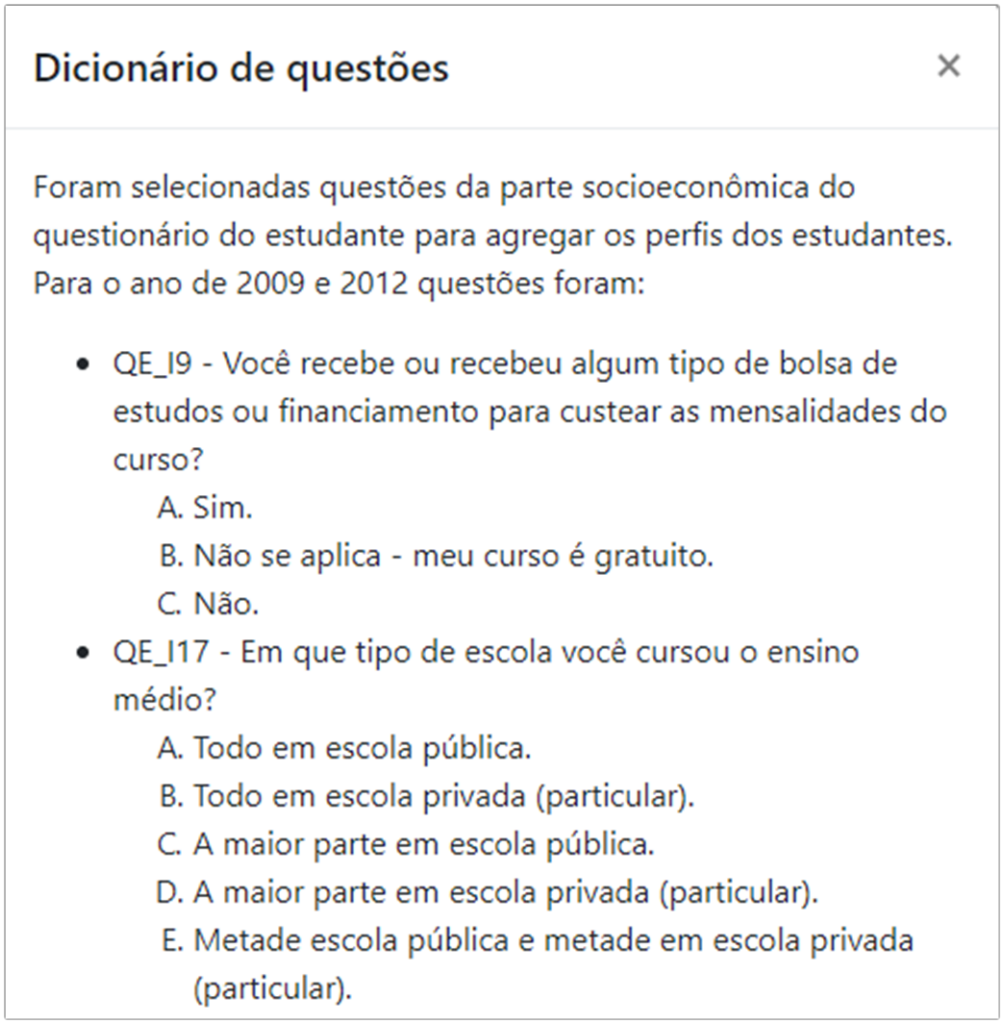
Figura 34 - Tabela de resultados de regras de associação



Ano	Curso	Regra	Total	Total em %
2009	Psicologia	Sudeste,Universidade => 4	1198	49.52%
2009	Psicologia	QE_17=B,QE_46=D => 4	808	33.40%
2009	Psicologia	PRIVADA,QE_46=D => 4	862	35.63%
2009	Psicologia	F,QE_46=D => 4	1074	44.40%
2009	Psicologia	QE_46=D,Universidade => 4	1207	49.90%
2009	Psicologia	PRIVADA,QE_17=B => 4	893	36.92%
2009	Psicologia	F,QE_17=B => 4	1221	50.48%
2009	Psicologia	QE_17=B,Universidade => 4	1351	55.85%

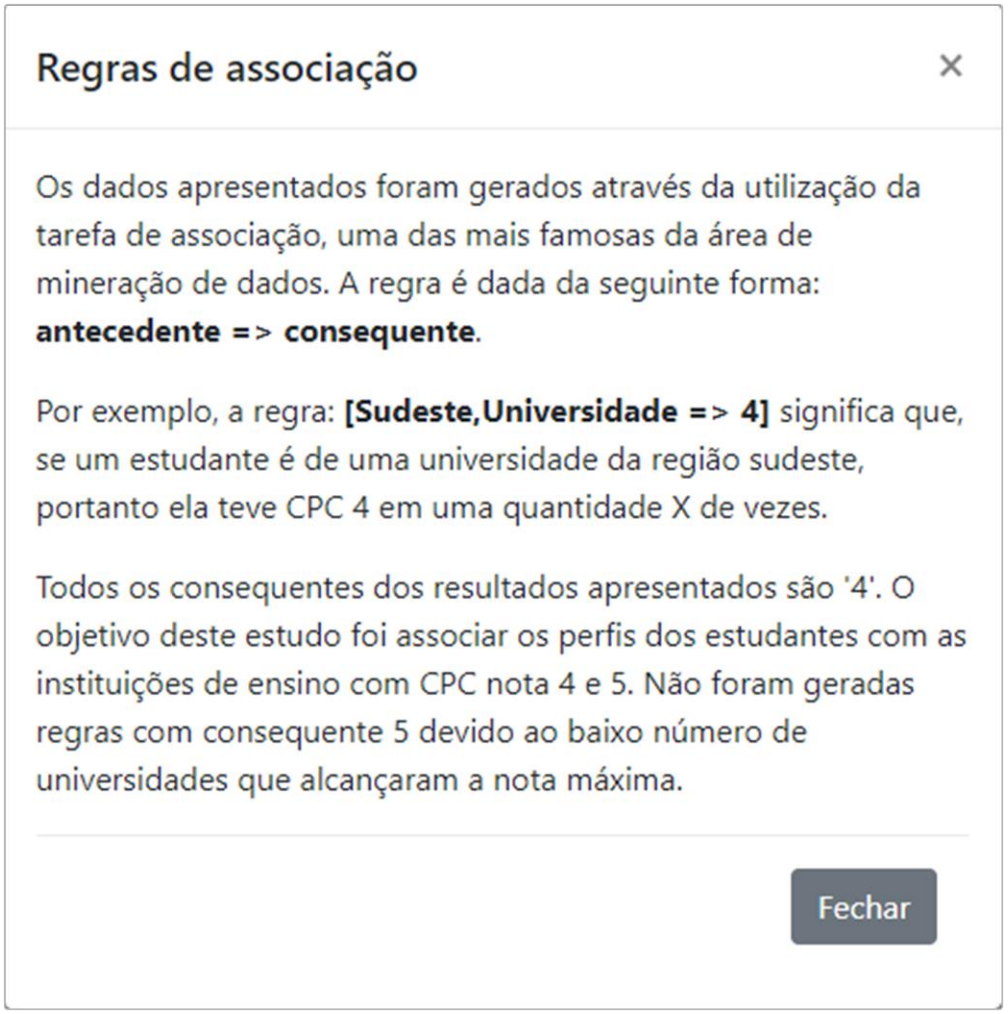
Além da exibição dos resultados através de gráficos, é possível visualizar os mesmos através de uma tabela na qual cada linha corresponde a uma regra gerada. É possível ver, na Figura 34, as informações contidas na tabela, que são: o ano correspondente aos dados analisados, o curso, a regra de associação e a quantidade de alunos englobados naquela regra, tanto em quantidade total quanto em porcentagem. Além dos resultados, também há um dicionário de questões e uma breve explicação de como são apresentadas as regras de associação.

Figura 35 - Dicionário de questões para os resultados de associação



O dicionário de questões apresentado na Figura 35 auxilia na interpretação dos resultados apresentados pelas regras de associação, mostrando cada questão respondida pelos alunos, bem como as alternativas de cada questão.

Figura 36 - Área de explicação das regras de associação



Também com intuito de auxiliar na interpretação dos resultados, a janela exibida na Figura 36 apresenta uma breve explicação de como funciona as regras de associação e o objetivo que levaram ao alcance dos resultados apresentados.

5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou a utilização de recursos de mineração de dados para transformação dos dados brutos do ENADE em conhecimentos passíveis de interpretação e auxílio para tomadas de decisão no meio acadêmico. O trabalho focou nos dados dos estudantes da área de Psicologia que participaram da prova do ENADE nos anos de 2009, 2012 e 2015. Na seção de referencial teórico foram abordados os recursos de mineração de dados e as características do CRISP-DM, modelo utilizado como base para a metodologia de desenvolvimento do trabalho.

Os conhecimentos descobertos durante o desenvolvimento foram dispostos em uma plataforma de apresentação de dados minerados, com foco em resultados de trabalhos realizados com os dados de cursos de graduação das instituições de ensino avaliadas no ENADE. Através do endereço https://enadedm.netlify.com é possível acessar a plataforma e visualizar as informações obtidas a partir de gráficos e da manipulação das informações utilizando filtros de busca. A plataforma dispõe informações descobertas com o uso das tarefas de mineração de dados associação e agrupamento, que foram utilizadas para obter diferentes formas de conhecimento nos dados trabalhados.

Com os resultados alcançados é possível acompanhar o desempenho dos estudantes da área de Psicologia separados pelas diferentes áreas que cobrem seus conteúdos. Também é possível verificar as características dos estudantes que fazem parte das instituições mais bem avaliadas no exame do ENADE nos três anos abordados no presente trabalho. Tais informações podem agregar conhecimento na tomada de decisão e melhor compreensão da situação da área de estudo no meio acadêmico.

Trabalhos futuros podem fornecer a mineração dos dados da área de Psicologia classificando-os pela matriz curricular geral do curso. Dados de outros cursos também podem ser explorados e adicionados na plataforma de apresentação de resultados. A criação de bibliotecas gráficas voltadas para apresentação de dados resultantes dos algoritmos de mineração de dados pode ajudar no melhor aproveitamento dos resultados obtidos, diminuindo a necessidade de maiores tratamentos dos dados para adequação a recursos gráficos não apropriados para a visualização de dados de mineração.

REFERÊNCIAS

AMO, Sandra de. **Técnicas de Mineração de Dados**. 2004. 43 f. Curso de Computação, Universidade Federal de Uberlândia, Uberlândia, 2004. Disponível em: <http://files.sistemas2012.webnode.com.br/200000095-bf367bfb43/Tecnicas%20de%20Mineração%20de%20Dados.pdf>. Acesso em: 07 set. 2018.

AZEVEDO, Ana Isabel Rojão Lourenço; SANTOS, Manuel Filipe. KDD, SEMMA and CRISP-DM: a parallel overview. **IADS-DM**, 2008. Disponível em: <http://recipp.ipp.pt/bitstream/10400.22/136/3/KDD-CRISP-SEMMA.pdf>. Acesso em: 01 nov. 2018.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de dados: Conceitos, tarefas, métodos e ferramentas.** Universidade Federal de Goiás (UFC), p. 1-29, 2009. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf>. Acesso em: 25 out. 2018.

CARVALHO, Luís Alfredo Vidal de. **A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração.** São Paulo: Érica, 2001. 234 p.

COSTA, Evandro; BAKER, Ryan S.J.d; AMORIM, Lucas; MAGALHÃES, Jonatas; MARINHO, Tarsis. Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. **Jornada de Atualização em Informática na Educação**, v. 1, n. 1, p. 1-29, 2012. Disponível em: <http://br-ie.org/pub/index.php/pie/article/view/2341/2096>. Acesso em: 25 nov. 2018.

DA COSTA CÔRTES, Sérgio; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. **Mineração de dados-funcionalidades, técnicas e abordagens.** PUC, 2002. Disponível em: <ftp://139.82.16.194/pub/docs/techreports/02_10_cortes.pdf>. Acesso em: 25 nov. 2018.

DA SILVA, Mabel Pereira; BOSCARIOLI, Clodis; PERES, Sarajane Marques. **Análise de logs da web por meio de técnicas de data mining.** 2003. Disponível em: <http://conged.deinfo.uepg.br/~iconged/Artigos/Artigo_03.pdf>. Acesso em: 07 set. 2018.

DE CASTRO, Armando Antonio Monteiro; DO PRADO, Pedro Paulo Leite. **Algoritmos para reconhecimento de padrões.** Revista Ciências Exatas, v. 8, n. 2002, 2001. Disponível em: <https://www.researchgate.net/profile/Pedro_Prado4/publication/277123850_Pattern_recognition_algorithms/links/55aa982b08ae481aa7fbc655.pdf>. Acesso em: 25 nov. 2018.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques.** 3. ed. Waltham: Elsevier, 2012. 673 p. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=pQws07tdpjoC&oi=fnd&pg=PP1&dq=Data+Mining:+Concepts+and+Techniques&ots=tzIzYZkzZ0&sig=9nf8GabG-Z3Z_X7KBjGB56gahug#v=onepage&q=Data%20Mining%3A%20Concepts%20and%20Techniques&f=false >. Acesso em: 30 maio 2019.

IBM BUSINESS ANALYTICS. **Guia do IBM SPSS Modeler CRISP-DM.** Rio de Janeiro, 2017. 52 p. Disponível em: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.1/br_po/ModelerCRISPDM.pdf>. Acesso em: 28 out. 2018.

INEP. **Enade**. 2015. Disponível em: <http://portal.inep.gov.br/enade>. Acesso em: 07 set. 2018.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados**. São Paulo: Novatec, 2015. 176 p. Disponível em: <https://www.researchgate.net/publication/282218981_Dados_Abertos_Conectados_em_Busca_da_Web_do_Conhecimento>. Acesso em: 22 set. 2018.

PIMENTEL, Edson P.; FRANÇA, Vilma F. De; OMAR, Nizam. **A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização.** In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2003. p. 495-504. Disponível em: <http://www.br-ie.org/pub/index.php/sbie/article/view/280/266>. Acesso em: 25 nov. 2018.

SAINI, Indu; SINGH, Dilbag; KHOSLA, Arun. **QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases.** Journal of advanced research, v. 4, n. 4, p. 331-344, 2013. Disponível em: <https://ac.els-cdn.com/S209012321200046X/1-s2.0-S209012321200046X-main.pdf?_tid=dc4a9b1f-20e1-4427-bb54-9599826ec1ab&acdnat=1540579260_3fb415673d7646c9216e6b4274a7fed0>. Acesso em: 26 out. 2018.

SILVA, Leandro A.; MORINO, Anderson Hideki; SATO, Thiago Massahiro Conti. **Prática de mineração de dados no exame nacional do ensino médio.** In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2014. p. 651. Disponível em: <http://www.br-ie.org/pub/index.php/wcbie/article/view/3289/2827>. Acesso em: 19 out. 2018.

VISTA, Nicolas Pastorio Boa; FIGUEIRÓ, Michele Ferraz; CHICON, Patricia Mariotto Mozzaquatro. **Técnicas de mineração de dados aplicadas aos microdados do ENADE para avaliar o desempenho dos acadêmicos do curso de Ciência da Computação no Rio Grande do Sul utilizando o software R.** I Seminário de Pesquisa Científica e Tecnológica, v. 1, n. 1, 2017. Disponível em: <http://www.revistaeletronica.unicruz.edu.br/index.php/revistaeletronica/article/view/5401/1138>. Acesso em: 19 out. 2018.

VOZNIKA, Fabricio; VIANA, Leonardo. **Data Mining Classification.** CSEP 521 - University of Washington. 2007. Disponível em: <https://courses.cs.washington.edu/courses/csep521/07wi/prj/leonardo_fabricio.pdf>. Acesso em: 25 out. 2018

WIRTH, Rüdiger; HIPP, Jochen. CRISP-DM: **Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining.** Citeseer, 2000. p. 29-39. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf>. Acesso em: 28 out. 2018.

APÊNDICES

APÊNDICE A – Questões e alternativas que agregaram os perfis dos estudantes de Psicologia para o ano de 2009 e 2012

- QE_I9: questão objetiva do questionário socioeconômico “Você recebe ou recebeu algum tipo de bolsa de estudos ou financiamento para custear as mensalidades do curso?”;
    - Alternativa A: Sim.
    - Alternativa B: Não se aplica - meu curso é gratuito.
    - Alternativa C: Não.
- QE_I17: questão objetiva do questionário socioeconômico “Em que tipo de escola você cursou o ensino médio?”;
    - Alternativa A: Todo em escola pública.
    - Alternativa B: Todo em escola privada (particular).
    - Alternativa C: A maior parte em escola pública.
    - Alternativa D: A maior parte em escola privada (particular).
    - Alternativa E: Metade em escola pública e metade em escola privada (particular).
- QE_I35: questão objetiva do questionário socioeconômico “Os conteúdos trabalhados pela maioria dos professores são coerentes com os que foram apresentados nos respectivos planos de ensino?”;
    - Alternativa A: Sim.
    - Alternativa B: Sim, somente em parte.
    - Alternativa C: Nenhum.
    - Alternativa D: Não sei responder.
- QE_I46: questão objetiva do questionário socioeconômico “Você participou de programas de iniciação científica? Como foi a contribuição para a sua formação?”;
    - Alternativa A: Sim, participei e teve grande contribuição.
    - Alternativa B: Sim, participei e teve pouca contribuição.
    - Alternativa C: Sim, participei e não percebi nenhuma contribuição.
    - Alternativa D: Não participei, mas a instituição oferece.
    - Alternativa E: A instituição não oferece esse tipo de programa.
- QE_I50: questão objetiva do questionário socioeconômico “Como você avalia o nível de exigência do curso?”.
    - Alternativa A: Deveria exigir muito mais.
    - Alternativa B: Deveria exigir um pouco mais.

- Alternativa C: Exige na medida certa.
- Alternativa D: Deveria exigir um pouco menos.
- Alternativa E: Deveria exigir muito menos.

APÊNDICE B – Questões e alternativas que agregaram os perfis dos estudantes de Psicologia para o ano de 2015

- QE_10: Qual alternativa a seguir melhor descreve sua situação de trabalho (exceto estágio ou bolsas)?
    - Alternativa A: Não estou trabalhando;
    - Alternativa B: Trabalho eventualmente;
    - Alternativa C: Trabalho até 20 horas semanais;
    - Alternativa D: Trabalho de 21 a 39 horas semanais;
    - Alternativa E: Trabalho 40 horas semanais ou mais.
- QE_11: Que tipo de bolsa de estudos ou financiamento do curso você recebeu para custear todas ou a maior parte das mensalidades? No caso de haver mais de uma opção, marcar apenas a bolsa de maior duração.
    - Alternativa A = Nenhum, pois meu curso é gratuito.
    - Alternativa B = Nenhum, embora meu curso não seja gratuito.
    - Alternativa C = ProUni integral.
    - Alternativa D = ProUni parcial, apenas.
    - Alternativa E = FIES, apenas.
    - Alternativa F = ProUni Parcial e FIES.
    - Alternativa G = Bolsa oferecida por governo estadual, distrital ou municipal.
    - Alternativa H = Bolsa oferecida pela própria instituição.
    - Alternativa I = Bolsa oferecida por outra entidade (empresa, ONG, outra).
    - Alternativa J = Financiamento oferecido pela própria instituição.
    - Alternativa K = Financiamento bancário.
- QE_14: Durante o curso de graduação você participou de programas e/ou atividades curriculares no exterior?
    - Alternativa A = Não participei.
    - Alternativa B = Sim, Programa Ciência sem Fronteiras.
    - Alternativa C = Sim, programa de intercâmbio financiado pelo Governo Federal (Marca; Brafitec; PLI; outro).
    - Alternativa D = Sim, programa de intercâmbio financiado pelo Governo Estadual.
    - Alternativa E = Sim, programa de intercâmbio da minha instituição.
    - Alternativa F = Sim, outro intercâmbio não institucional.

- QE_15: Seu ingresso no curso de graduação se deu por meio de políticas de ação afirmativa ou inclusão social?
    - Alternativa A = Não.
    - Alternativa B = Sim, por critério étnico-racial.
    - Alternativa C = Sim, por critério de renda.
    - Alternativa D = Sim, por ter estudado em escola pública ou particular com bolsa de estudos.
    - Alternativa E = Sim, por sistema que combina dois ou mais critérios anteriores.
    - Alternativa F = Sim, por sistema diferente dos anteriores.
- QE_18: Qual modalidade de ensino médio você concluiu?
    - Alternativa A = Todo em escola pública.
    - Alternativa B = Todo em escola privada (particular).
    - Alternativa C = Todo no exterior.
    - Alternativa D = A maior parte em escola pública.
    - Alternativa E = A maior parte em escola privada (particular).
    - Alternativa F = Parte no Brasil e parte no exterior.