

**MINERAÇÃO DE DADOS: UM ESTUDO PARA IDENTIFICAÇÃO DO PERFIL
DOCENTE DAS IES COM CONCEITO 3 OU SUPERIOR NO IGC AVALIADO EM
2016 NO ESTADO DO ESPÍRITO SANTO****André Bessa da Silva**

Faculdade Multivix Cachoeiro de Itapemirim – Campus 1
Rua Moreira, 29, Independência. Cachoeiro de Itapemirim/ES | CEP 29306-320.
andrebessax@gmail.com

Denilton Macário de Paula

Instituto Federal do Espírito Santo – Campus Vianna
Rodovia BR-101, Km 12, S/N Universal, Viana/ES | CEP: 29135-000.
denilton.macario@ifes.edu.br

Geórgia Regina Rodrigues Gomes

Universidade Federal Fluminense INFES/UFF
Avenida João Jasbick, S/N, Aeroporto, Santo Antônio de Pádua/RJ | CEP 28470-000.
georgiargomes@gmail.com

RESUMO

Neste trabalho de pesquisa o objetivo é aplicar técnicas de mineração de dados para extrair informações relevantes da base de dados do censo superior de 2016, disponibilizados pelo INEP, sendo utilizada como metodologia uma pesquisa de revisão bibliográfica para conceituação de termos e algoritmos de mineração dados a serem utilizados na pesquisa, bem como discernimento dos processos e indicadores que o MEC leva em consideração para avaliar uma instituição de ensino superior no Brasil. Além disso, também é aplicado o tratamento dos dados anterior a fase de mineração e aplicação do processo do KDD. A pesquisa resulta na criação de clusters de informações, que agrupam os dados da mineração em 7 grandes grupos, além de detalhar algumas ramificações contidas na árvore de decisão criada com base no algoritmo J48. Com este trabalho, podemos concluir que a mineração de dados se apresenta como uma ferramenta poderosa para extração de conhecimento, que nos possibilitou traçar perfis dos docentes capixabas das instituições com IGC 3 ou superior.

Palavra-chave: Algoritmos, Clusters, Ensino Superior Mineração de dados, KDD, J48.

ABSTRACT

In this research work the use of data mining techniques to extract relevant data from the 2016 superior database provided by INEP, having been used as a bibliographic research source for conceptualizing data mining terms and algorithms. The MEC takes into consideration the choice of a higher education institution in Brazil, as well as the treatment of data prior to a mining phase and application of the KDD process. The research resulted in the creation of information sets, which group the mining data into 7 large groups, and have some branches contained in the analysis area based on the J48 algorithm. With this work, we can conclude that data mining is a powerful tool for knowledge extraction, which can be found in the curriculum of Espírito Santo teachers from institutions with IGC 3 or higher.

Keywords: Algorithms, Clusters, Higher Education Data Mining, KDD, J48.

Como Citar:

SILVA, André Bessa; De PAULA, Denilton Macário; GOMES, Geórgia Regina Rodrigues. Mineração de dados: um estudo para identificação do perfil docente das ies com conceito 3 ou superior no igc avaliado em 2016 no estado do espírito santo. *In: SIMPÓSIO DE PESQUISA OPERACIONAL E LOGÍSTICA DA MARINHA*, 19., 2019, Rio de Janeiro, RJ. **Anais** [...]. Rio de Janeiro: Centro de Análises de Sistemas Navais, 2019.

1. INTRODUÇÃO

Este artigo tem por objetivo apresentar uma análise do perfil dos docentes em instituições de ensino superior no estado do Espírito Santo, levando em consideração o índice geral de cursos (IGC) com índice satisfatório ou superior, este índice hoje se constitui de uma métrica que afere qualidade do curso superior em instituições públicas e privadas. Para a confecção do estudo foram analisados os Microdados do Censo 2016, que está disponível para consulta pública no site do INEP, e nesta base de dados foram aplicados técnicas de mineração de dados usando o software Weka 3.8.3.

O Ministério da Educação (MEC) possui hoje uma série de instrumentos que ele utiliza para a avaliação do Ensino Superior no Brasil, hoje grande parte das diretrizes de avaliação são reguladas pelo SINAES, sistema nacional de avaliação do Ensino Superior, que congrega um conjunto de conceitos, dentre os quais destacamos o Índice Geral de Cursos (IGC), que hoje assume grande importância para as instituições de Ensino Superior de modo geral, mas em especial as faculdades, centros universitários e universidades privadas.

Os conceitos gerados pelas avaliações do MEC e de institutos abarcados pelo ministério para avaliar o ensino superior, podem fazer uma diferença significativa na opção de escolhas de cursos a serem cursados nestas instituições, além de ser um importante chamariz para o marketing intra e extra institucional, por permitir a criação de rankings que indicam o grau de qualidade dos cursos de determinada instituição (Ristoff e Giolo, 2006).

Como Cunha et. Al (2003) expõe no cenário atual das políticas de avaliação do ensino superior, o índice alcançado, assume um papel importante para conferir ou associar um alto grau de qualidade, nos casos de conceitos positivos, este passa a ser um grande trunfo para as campanhas de marketing da instituição e do curso, e quando mal avaliado é um agente impulsionador dentro da instituição para buscar melhorias para boas conceituações futuras.

2. ENSINO SUPERIOR NO BRASIL

No Brasil o ensino superior é ofertado por instituições que se classificam administrativamente como universidades, centros universitários, faculdades, institutos federais e a regulação e avaliação são incumbências do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e pelo Ministério da Educação (MEC).

Por meio de avaliações como o ENADE o MEC avalia e acompanha o processo de ensino e aprendizagem e o desempenho dos discentes em relação ao curso e disciplinas cursadas nas mais diversas graduações.

O Exame Nacional de Desempenho dos Estudantes (Enade) é um dos pilares da avaliação do Sistema Nacional de Avaliação da Educação Superior (Sinaes), criado pela Lei nº10.861, de 14 de abril de 2004. O Sinaes é composto, ainda, pelos processos de Avaliação de Cursos de Graduação e de Avaliação Institucional que, junto com o Enade, formam um ‘tripé’ avaliativo que permite conhecer em profundidade o modo de funcionamento e a qualidade dos cursos e instituições de educação superior (IES) de todo o Brasil. (INEP, 2016).

Como consta no SINAES vol. 5 do INEP (2015), o IGC é um indicativo da qualidade dos cursos oferecidos pelas instituições de ensino superior levando em consideração a conceituação obtida pela instituição no ENADE. O IGC considera além dos aspectos sociais e desempenho discente, também a formação docente e infraestrutura da instituição de ensino superior.

Como apontado na Figura 1, os resultados do IGC 2016, mostram que grande parte das instituições brasileiras encontram-se na faixa C (nota 3), que é um índice satisfatório, mas qual o perfil dos docentes das instituições D e E? Este estudo utiliza esta indagação para o escopo do estado do Espírito Santo, para ajudar a traçar o perfil destes docentes.

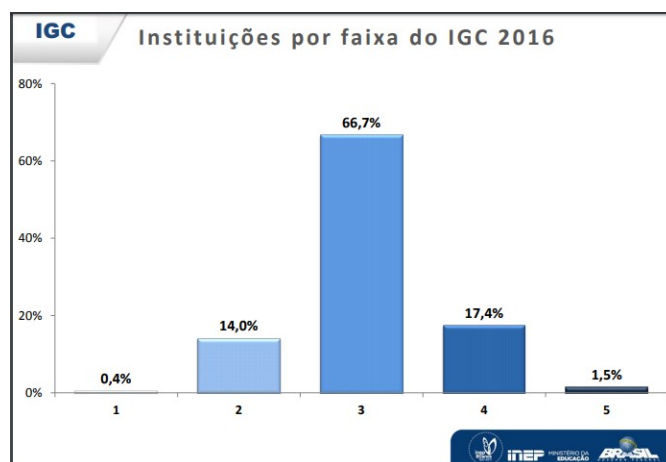


Figura 1 – Resultado do IGC 2016. Fonte: INEP. Disponível em <
http://download.inep.gov.br/educacao_superior/igc_cpc/2016/apresentacao_resultados_CPC_e_IGC_2016.pdf>

3. MINERAÇÃO DE DADOS

A coleta de informação de caráter relevante, hoje assume grande importância independente da área de estudo ou conhecimento e se mostra cada vez mais necessária para empresas e organizações para ações estratégicas e como suporte a tomada de decisão. Portanto se faz necessário o conhecimento de técnicas e ferramentas que a isso propicie.

Soma-se a isso o grande volume de informação produzido, sobretudo nas últimas décadas no que se refere a implementação de enormes bases de dados transacionais, inteligência computacional e das redes de telecomunicações, neste cenário a mineração de dados pode ser usada como uma excelente ferramenta para busca de informações úteis nesta gigantesca massa de dados, possibilitando aos tomadores de decisão, a melhor ação a ser tomada em condições de certeza limitada (SOCZEK e ORLOVSKI, 2014).

De acordo com Cardoso e Machado (2008), dentro da área de banco de dados, a Mineração de dados, é uma disciplina que vai propor e apresentar técnicas e algoritmos com a finalidade de extrair informações relevantes de base de dados densamente povoadas. Ainda

segundo os autores, na mineração de dados, pode-se de maneira automatizada encontrar informações que estão “escondidas” nos dados, isso permite uma ação rápida pelo tomador de decisão com base nos dados encontrados.

Importante destacar que a mineração de dados é vista como parte de um processo maior de descoberta de conhecimento em bases de dados denominada Knowledge Discovery In Databases ou simplesmente KDD, há na literatura quem considere Data Mining e KDD sinônimos (FAYYAD et. al., 1996).

Tradicionalmente, ainda se faz uso de técnicas manuais para a coleta de dados, para a partir deste ponto, iniciar o processamento das informações por especialistas que decodificarão os dados e informação em uma série de relatórios a serem entregues aos tomadores de decisão, mas devido aos avanços em tecnologias, que hoje, geram um grande volume de dados, a coleta e análise manual, torna-se impraticável, e é neste momento que a mineração de dados torna-se um interessante ferramenta na busca e análise de dados. (CAMILO e CARLOS, 2009).

Remontando aos objetivos deste artigo, será realizada a mineração de dados em uma base de dados do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) do ano de 2016 com as informações dos docentes, colocando o foco da mineração na identificação do perfil docente das instituições de ensino superior no estado do Espírito Santo para instituições com IGC satisfatório ou superior, apresentando o processo de seleção e transformação dos dados usando o KDD (Knowledge Discovery In Databases).

4. KDD (KNOWLEDGE DISCOVERY IN DATABASES)

No processo de mineração de dados o KDD é o processo que visa encontrar os padrões dentro das volumosas bases de dados que temos, sendo caracterizada como o núcleo do processo de mineração que envolve seus algoritmos de exploração de dados e modelagem de padrões até então desconhecidos na base de dados (MAIMON, 2009).

De acordo com Vianna (2010), o KDD é o conjunto de tarefas que permite a busca de conhecimento em enormes bancos de dados, neste processo de investigação se dará através do levantamento, seleção dos dados, o pré-processamento e pôr fim a mineração, que permitirá a transformação dos dados extraídos em conhecimento útil à tomada de decisão. Na Figura 2 é apresentado o processo do KDD.

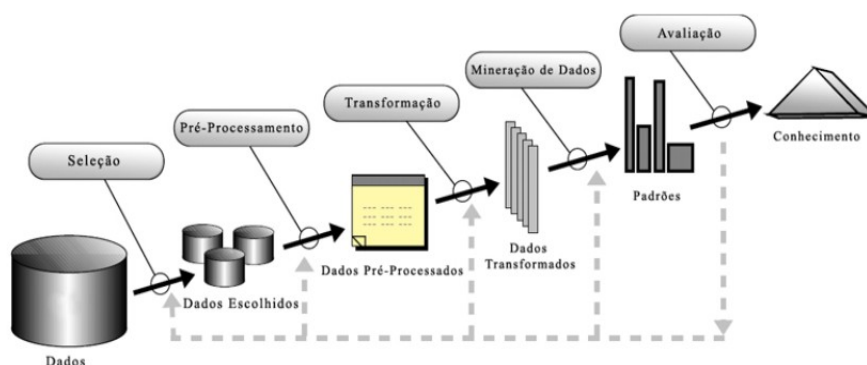


Figura 2 - Processo de Mineração de Dados com base no KDD

FONTE: CAMILO E CARLOS (2009).

5. METODOLOGIA

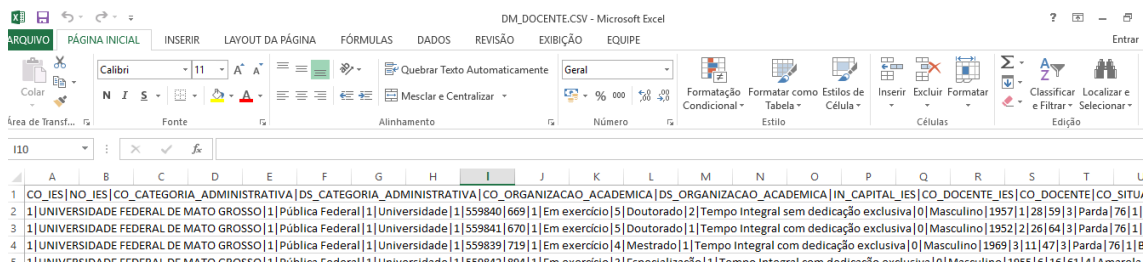


Figura 4 – Recorte parcial da base de dados com informações do docente da IES antes do processo de tratamento da base.

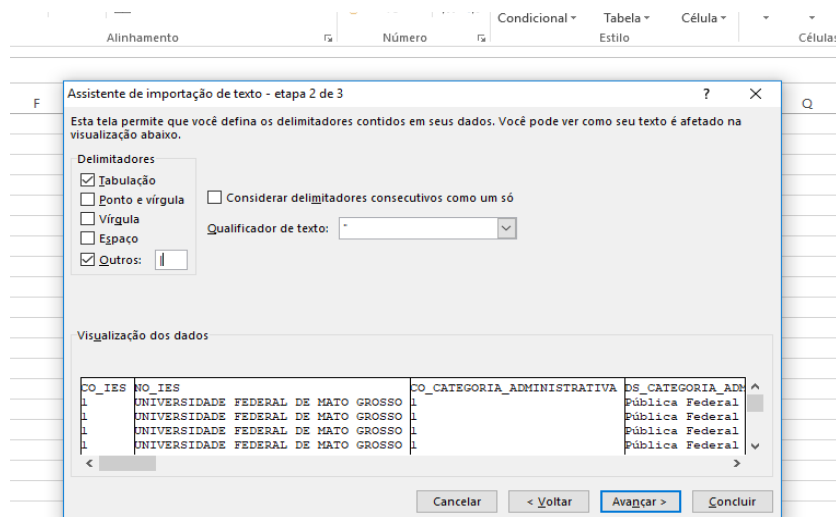


Figura 5 – Ferramenta Microsoft Excel para abertura do arquivo de microdados do censo 2016.

Nesta base foi realizado um processo de extração dos dados de instituição e docentes do Ensino Superior do estado do Espírito Santo, formando uma base de dados 7234 registros. Vale ressaltar que foi realizada uma limpeza e tratativa nos dados a fim de eliminar registros com informações incompletas, ou com valores nulos, pois sem isso aplicar o processo de mineração poderia causar inconsistências de informações geradas neste estudo.

Um detalhe importante sobre os dados do censo 2016, é que o mesmo não informa os conceitos do IGC das instituições de Ensino Superior, sendo necessária a inclusão desta informação em uma coluna em nossa base de dados, como apresentada na Figura 6.

The screenshot shows the Microsoft Excel interface with the file 'BASE DE DADOS - CENSO MICRODADOS INEP 2016.xlsx'. The data table includes a new column 'IGC_DESC' which categorizes institutions by IGC.

1	DS_ESCOLARIDADE_DOCENTE	DS_REGIME_TRABALHO	DS_REG_TRAB	DS_SEXO_DOCENTE	IDADE_FAT	IDADE	ICG	IGC_DESC	DS_COR_RACA_DOCENTE
2	Mestrado	Tempo Integral sem dedicação exclusiva	DE SEM	Feminino	45 a 54	50	3	C	branca
3	Mestrado	Tempo Parcial	PARCIAL	Masculino	36 a 44	40	3	C	branca
4	Especialização	Horista	HORISTA	Feminino	> 54	55	3	C	branca
5	Especialização	Tempo Parcial	PARCIAL	Feminino	> 54	64	3	C	branca
6	Especialização	Tempo Parcial	PARCIAL	Feminino	> 54	58	3	C	branca
7	Mestrado	Tempo Parcial	PARCIAL	Feminino	45 a 54	53	3	C	branca
8	Especialização	Tempo Parcial	PARCIAL	Masculino	45 a 54	51	3	C	branca
9	Mestrado	Tempo Parcial	PARCIAL	Feminino	36 a 44	42	3	C	branca
10	Doutorado	Tempo Parcial	PARCIAL	Masculino	> 54	63	3	C	branca
11	Especialização	Tempo Parcial	PARCIAL	Feminino	> 54	62	3	C	branca
12	Especialização	Tempo Parcial	PARCIAL	Feminino	> 54	61	3	C	branca
13	Especialização	Tempo Integral sem dedicação exclusiva	DE SEM	Feminino	> 54	60	3	C	branca
14	Mestrado	Tempo Parcial	PARCIAL	Feminino	45 a 54	52	3	C	branca
15	Especialização	Tempo Parcial	PARCIAL	Feminino	45 a 54	52	3	C	branca
16	Especialização	Tempo Parcial	PARCIAL	Feminino	> 54	67	3	C	branca
17	Especialização	Horista	HORISTA	Feminino	45 a 54	52	3	C	branca
18	Especialização	Tempo Parcial	PARCIAL	Masculino	45 a 54	47	3	C	branca
19	Especialização	Horista	HORISTA	Masculino	45 a 54	51	3	C	branca

Figura 6 – Recorte parcial da base de dados com a adição da coluna que conceitua a instituição com o IGC.

Após a tratativa da base de dados para aplicarmos a mineração de dados utilizando os dados do censo 2016 para identificação dos perfis docentes das instituições IGC 3 ou superior no estado do Espírito Santo foram destacados 9 atributos chaves, conforme a quadro 1, tomando como referência o trabalho Batista et Al (2015).

Quadro 1: Atributos da base de dados a ser minerada.

ATRIBUTO	DESCRIÇÃO	VALOR
DS_CATEGORIA_ADMINISTRATIVA	Categoria Administrativa	PUBLICA FEDERAL
		PUBLICA ESTADUAL
		PUBLICA MUNICIPAL
		PRIVADA
		PRIVADA
		ESPECIAL
DS_ORGANIZACAO_ACADEMICA	Organização acadêmica	UNIVERSIDADE
		CENTRO UNIVERSITARIO
		FACULDADE
		INSTITUTO FEDERAL
LOCALIZACAO	Localização da IES	CAPITAL
		INTERIOR
DS_ESCOLARIDADE_DOCENTE	Titulação Acadêmica	GRADUAÇÃO
		ESPECIALIZAÇÃO
		MESTRADO
		DOUTORADO
DS_REG_TRAB	Regime de Trabalho	DE SEM (Integral sem dedicação exclusiva)
		PARCIAL
		HORISTA
		DE (Integral com Dedicação exclusiva)
DS_SEXO_DOCENTE	Sexo Docente	MASCULINO
		FEMININO
IDADE	Faixa Etária	20 Á +64
IGC_DESC	Conceito do IGC do MEC	C = 3
		D = 4
		E = 5
DS_COR_RACA_DOCENTE	Raça declarada	Docente não quis declarar cor/raça
		Parda
		Branca
		Preta
		Índio
		Amarela

Para realizarmos a mineração nos dados do censo docente 2016, dentre os diversos algoritmos disponíveis na ferramenta Weka 3.8.3 foi utilizado para este estudo à

classificação dos dados com a Clusterização e a análise de árvores de decisão usando o algoritmo J48.

7. PROCESSAMENTO

No desenvolvimento deste trabalho, optou-se por usar a ferramenta de mineração de dados Weka 3.8.3 desenvolvida em linguagem Java pela Universidade Waikato em Hamilton, na Nova Zelândia, Figura 7. Nesta ferramenta são agrupados algoritmos e técnicas de mineração de dados como J48, Apriori, Clusterização além de contar com uma série de processo que auxiliam no pré-processamento da base de dados, que pode ser oriunda de um arquivos texto ou separado por vírgulas (.csv) ou ainda pode-se conectar diretamente em bases de dados do tipo relacional como um banco Oracle, Mysql ou PostgreSQL.

8. TAREFA DE CLUSTERIZAÇÃO

Para a aplicação da mineração nos dados do INEP para o censo docente de 2016, os dados foram devidamente tratados anteriormente, um dos algoritmos usados foi o de Clusterização presente na ferramenta Weka 3.8.3.

De acordo com Jain et. al, (1999), a Clusterização é um tipo de classificação de dados não supervisionados, onde se formam os denominados Cluster ou simplesmente grupamentos. Sendo assim, podemos definir de uma maneira simples que a Clusterização é um algoritmo de agrupamento de dados multivariados que utiliza métodos numéricos, objetivando agrupar de maneira automática uma série de dados n em uns grupos ou clusters k de acordo com sua similaridade.

Portanto a idéia central do algoritmo é agrupar em mesmo cluster elementos com alto grau de similaridade (mesmo padrão), mas que devem ser diferentes de elementos de outros clusters, ou seja, dissimilares (CASSIANO, 2014).

Para Cassiano (2014) a grande vantagem de se usar a Clusterização é que ao agruparmos dados que sejam similares podemos descrever de forma mais clara e eficiente características de cada grupo de dados identificado. Fornecendo assim uma maior compreensão dos dados originais além de permitir criarmos esquemas de classificação e descobrir relações realmente interessantes entre os atributos dos dados de uma forma que se possam extrair informações, pelas quais usando as técnicas tradicionais não conseguiríamos alcançar.

Na confecção deste trabalho, para utilizarmos a Clusterização foi parametrizada a ferramenta para criação de 7 clusters de dados usando a técnica do “Simple K-means”, usando como função de distância um método Euclidiano.

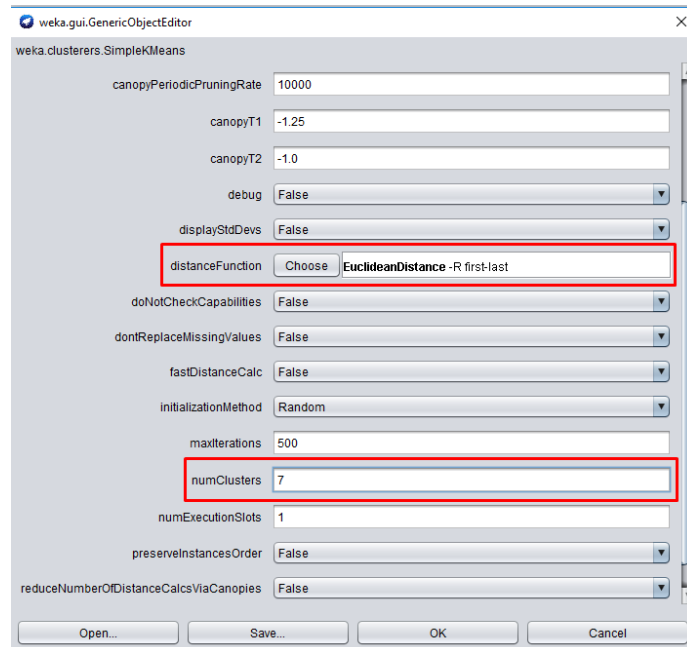


Figura 7 – Parametrização do cluster no Weka 3.8.3.

Monteiro e Silva et. Al (2001), afirmam que o K-means ou K-médias consiste em um método no qual um conjunto de valores iniciais n são agrupados como estimativas iniciais temporária das médias dos k clusters, onde k é o número de clusters a ser definido pelo usuário. A partir desse ponto os dados são agrupados de maneira a juntar os mais similares e depois comparados com os mais “distantes” e com os outros Clusters formados.

1 (1015,0)	2 (935,0)
Privada sem fins lucrativos	Privada sem fins lucrativos
Faculdade	Faculdade
CAPITAL	INTERIOR
Mestrado	Mestrado
PARCIAL	PARCIAL
Feminino	Feminino
45 a 54	36 a 44
D	C
Branca	Branca

Figura 8 – Recorte do resultado da execução da Clusterização utilizando o Weka 3.8.3, apresentação de 2 dos 7 clusters criados.

9. TAREFA DE CLASSIFICAÇÃO

O Algoritmo J48 possibilita a criação de modelos de árvore de decisão, nesta técnica o algoritmo gera árvores decisórias onde cada nó avalia a significância de cada atributo da

base de dados. O J48 é uma reescrita em linguagem de programação JAVA do algoritmo C45 escrito inicialmente em linguagem C (QUINLAN, 1993).

Na computação árvore s são estruturas de dados formadas por um conjunto de nós, que armazenam informações. Na hierarquia, toda árvore possui um nó raiz, que representa o mais alto nível da árvore, que é também seu ponto de partida para ligações entre os demais elementos da árvore s, sendo que um nó que não possui desmembramento ou filhos é chamado de nó folha (CAMPOS, 2017).

Como explica Campos (2017), que em uma árvore de decisão, toda decisão tomada, parte ou se encaminha do nó raiz até os nós folhas da árvore gerada.

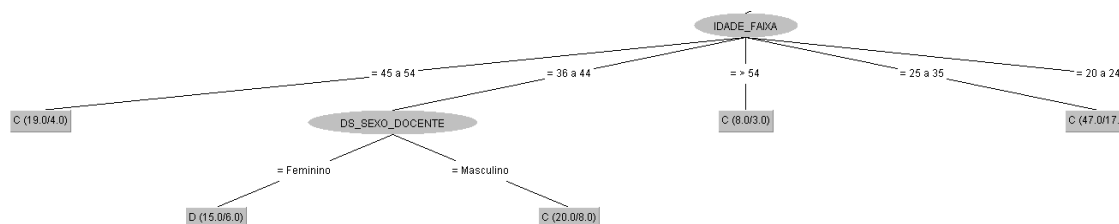


Figura 9– Exemplo de um nó folha em uma árvore. Imagem retirada da análise no Weka 3.8.3

Diante do exposto sobre árvore s de decisão, para utilização da mesma no Weka 3.8.3 foi parametrizado um fator de confiança de 0,25. Executado tendo como o nó folha o atributo IGC_DESC como pode ser analisado nas Figuras 10 respectivamente.

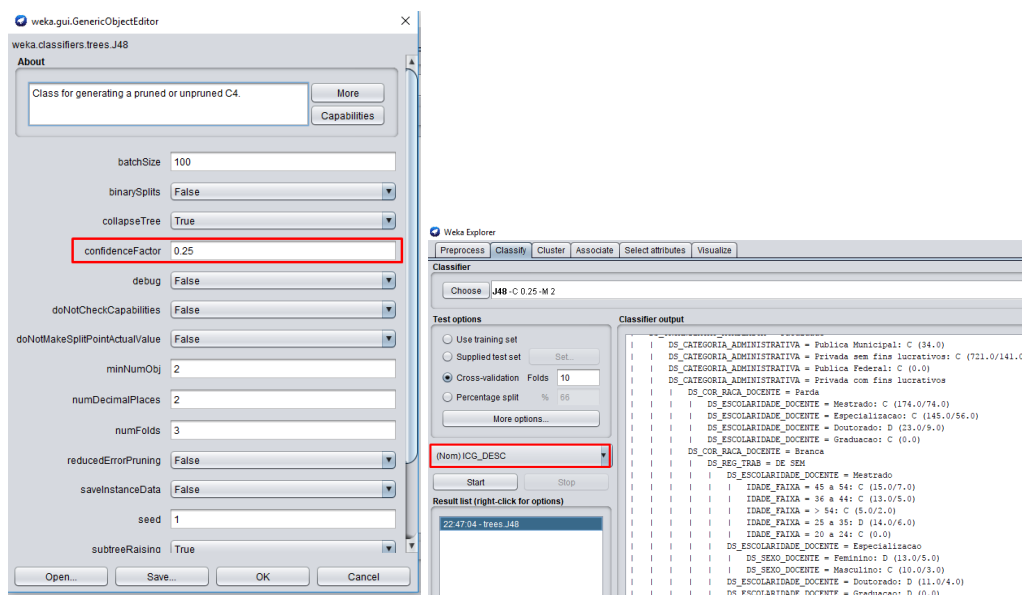


Figura 10 – Recorte da parametrização e resultado da execução do algoritmo J48 no Weka 3.8.3.

10. RESULTADOS E DISCUSSÕES

Utilizando como ferramenta o software Weka 3.8.3, foi carregado para o aplicativo uma base de dados no formato arquivo de texto separado por vírgula (.csv), base está que passou por um pré-processamento dos dados no software Microsoft Excel. Nesta base foram aplicados dois algoritmos de mineração a Clusterização para agrupamento de dados e o J48 para criação de uma árvore de decisão para análise dos ramos mais interessantes gerados, tendo como nó folha a informação do IGC da instituição vinculado ao docente.

Neste processo de mineração de dados foram gerados como resultados 7 clusters com os atributos definidos na tabela de atributos conforme explicitado na seção PRÉ-PROCESSAMENTO DE DADOS. Com 7 clusters agrupamos os dados de maneira simplificado e já podemos inferir conhecimento sobre a base de dados.

Na Figura 11 ficam ilustrados alguns dos clusters gerados no processo de mineração de dados dentro do Weka 3.8.3.

3 (1866.0)	4 (683.0)	5 (399.0)	6 (404.0)
Publica Federal	Publica Federal	Privada com fins lucrativos	Publica Federal
Universidade Instituto Federal de Educacao Ciencia e Tecnologia	Universidade	Faculdade	Universidade
CAPITAL	CAPITAL	CAPITAL	CAPITAL
Doutorado	Mestrado	Mestrado	Mestrado
DE	DE	HORISTA	DE SEM
Feminino	Masculino	Feminino	Masculino
36 a 44	25 a 35	36 a 44	25 a 35
D	D	D	D
Branca	Branca	Branca	Branca

Figura 11 – Recorte parcial dos clusters gerados no weka 3.8.3. Clusters de 3 a 6.

No primeiro cluster gerado, número 0, foi identificado uns grupos de docentes que atuam em instituições PRIVADAS COM FINS LUCRATIVOS, classificada como FACULDADE no INTERIOR do Espírito Santo em sua maioria com título de ESPECIALIZAÇÃO são do sexo MASCULINO entre 25 a 35 anos e com regime de trabalho PARCIAL, se autodeclaram BRANCOS e a instituição é IGC C.

No cluster número 1 outro grande agrupamento evidencia que instituições PRIVADAS SEM FINS LUCRATIVOS classificados como FACULDADE e localizados na CAPITAL, os docentes em sua maioria tem a titulação de MESTRADO, dedicação PARCIAL são do sexo FEMININO de 45 a 54 anos, são BRANCAS e a instituição é IGC D.

O terceiro cluster, número 2, agrupa as instituições FACULDADES PRIVADAS SEM FINS LUCRATIVA no INTERIOR do estado, os docentes têm MESTRADO como titulação e um regime PARCIAL de trabalho, são do sexo FEMININO com idades de 35 a 45 anos, brancas a instituição é IGC C.

Analisando o quarto cluster, número 3, é possível identificar um agrupamento onde instituições PUBLICA FEDERAL, denominadas como UNIVERSIDADE que se encontram

na CAPITAL, a titulação dos docentes é DOUTORADO que possuem dedicação exclusiva (DE), do sexo FEMININO de 36 a 44 anos, BRANCAS as instituições são IGC D.

Na análise do quinto cluster, número 4, as instituições PUBLICA FEDERAL, classificadas como INSTITUTO FEDERAL na CAPITAL tem docentes com MESTRADO com dedicação exclusiva (DE), do sexo MASCULINO com idades de 25 a 35 BRANCOS e o IGC do instituto é D.

No Sexto cluster o agrupamento reuniu instituições FACULDADES COM FINS LUCRATIVOS localizadas na CAPITAL, possuem docentes HORISTAS com titulação de MESTRADO são do sexo FEMININO com idades de 36 a 44 anos BRANCAS com IGC da instituição sendo D.

No último cluster criado, número 6, nas instituições PUBLICA FEDERAL classificadas como UNIVERSIDADES na CAPITAL os docentes com MESTRADO sem dedicação exclusiva (DE SEM) são do sexo MASCULINO, declaram-se BRANCOS com idades de 25 a 35 anos e a instituição é IGC D.

Na execução do algoritmo de árvore de decisão J48, foi gerada uma série de ramos que possibilitam a análise de um recorte interessante de dados para a análise na figura 12 e 13 ilustramos partes dos ramos da árvore gerada.

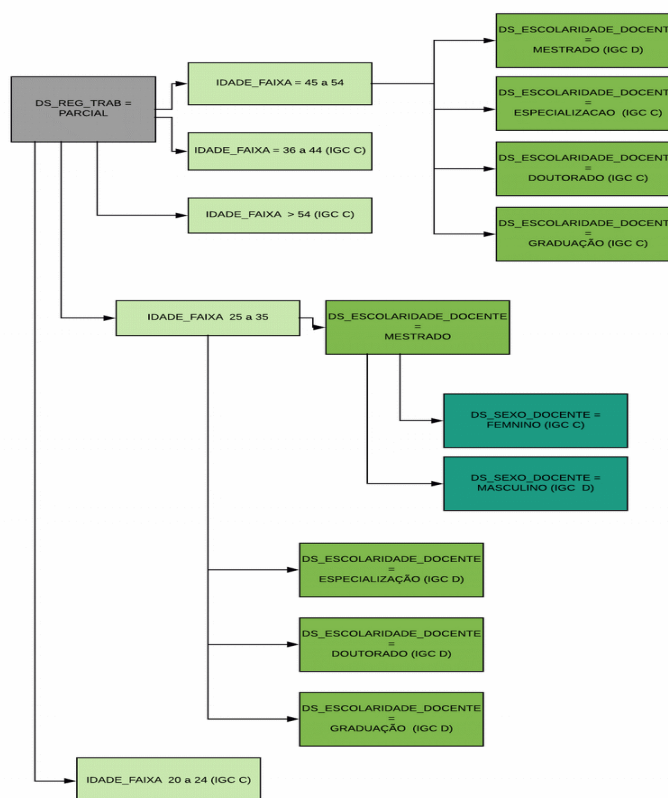


Figura 12 – Recorte parcial da árvore J48 com o atributo DS_REG_TRAB com o valor PARCIAL.

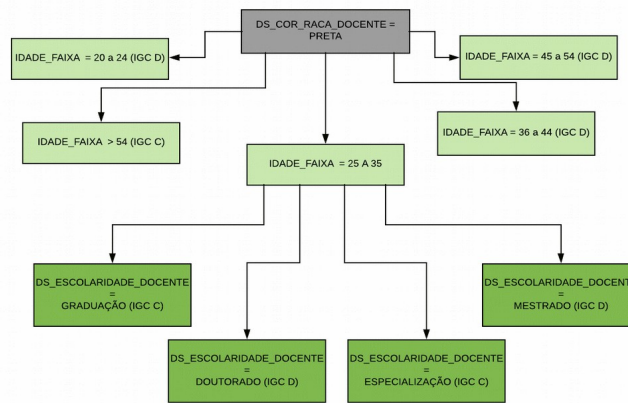


Figura 13 – Recorte parcial da árvore J48 com o atributo DS_COR_RACA_DOCENTE com o valor PRETO.

Na criação da árvore de decisão na ferramenta Weka, podemos analisar o corte da figura 16, onde a partir do nó DS_REG_TRAB = PARCIAL são criados ramos para os valores das faixas etárias (IDADE_FAIXA) onde para a faixa de 45 a 54 anos existem dados sobre a titulação docente e a classificação da instituição, onde: Mestrado (IGC D), Doutorado (IGC C), Especialização (IGC C) e Graduação (IGC C).

Para um novo corte na árvore de decisão J48, analisamos o Nó DS_COR_RACA_DOCENTE = PRETA, na qual é possível ramificar a IDADE_FAIXA de 25 a 35 com a titulação docente e a classificação da instituição, sendo: Mestrado (IGC D), Doutorado (IGC D), Especialização (IGC C) e Graduação (IGC C).

Abaixo na Figura 14e 15 é apresentado o Log da árvore decisão J48 após a execução do algoritmo na ferramenta Weka 3.8.3.

```
DS_REG_TRAB = PARCIAL
| IDADE_FAIXA = 45 a 54
| | DS_ESCOLARIDADE_DOCENTE = Mestrado: D (38.0/18.0)
| | DS_ESCOLARIDADE_DOCENTE = Especializacao: C (22.0/7.0)
| | DS_ESCOLARIDADE_DOCENTE = Doutorado: C (13.0/6.0)
| | DS_ESCOLARIDADE_DOCENTE = Graduacao: C (0.0)
| IDADE_FAIXA = 36 a 44: C (114.0/43.0)
| IDADE_FAIXA = > 54: C (41.0/10.0)
| IDADE_FAIXA = 25 a 35
| | DS_ESCOLARIDADE_DOCENTE = Mestrado
| | | DS_SEXO_DOCENTE = Feminino: C (51.0/24.0)
| | | DS_SEXO_DOCENTE = Masculino: D (52.0/22.0)
| | DS_ESCOLARIDADE_DOCENTE = Especializacao: D (66.0/31.0)
| | DS_ESCOLARIDADE_DOCENTE = Doutorado: D (12.0/5.0)
| | DS_ESCOLARIDADE_DOCENTE = Graduacao: D (0.0)
| IDADE_FAIXA = 20 a 24: C (3.0)
```

Figura 14 – Recorte parcial da árvore J48 com o atributo DS_REG_TRAB com o valor PARCIAL
Imagem de Log da ferramenta Weka 3.8.3.

```

DS_COR_RACA_DOCENTE = Preta
| IDADE_FAIXA = 45 a 54: D (4.0/1.0)
| IDADE_FAIXA = 36 a 44: D (10.0/2.0)
| IDADE_FAIXA = > 54: C (4.0)
| IDADE_FAIXA = 25 a 35
| | DS_ESCOLARIDADE_DOCENTE = Mestrado: D (7.0/2.0)
| | DS_ESCOLARIDADE_DOCENTE = Especializacao: C (13.0/4.0)
| | DS_ESCOLARIDADE_DOCENTE = Doutorado: D (4.0/2.0)
| | DS_ESCOLARIDADE_DOCENTE = Graduacao: C (0.0)
| IDADE_FAIXA = 20 a 24: D (0.0)

```

Figura 15 – Recorte parcial da árvore J48 com o atributo DS_COR_RACA_DOCENTE com o valor PRETA. Imagem de LOG na ferramenta Weka 3.8.3

11. RESULTADOS E DISCUSSÕES

Utilizar o processo de mineração de dados nos dados do censo 2016 do ensino superior para traçarmos o perfil do docente capixaba com índice satisfatório no IGC com o auxílio de rotinas inclusas do KDD e com os algoritmos de mineração presentes no software Weka na sua versão 3.8.3 se mostrou muito eficiente. Conseguindo criar Cluster com dados relevantes para uma visão macro das informações da base e em uma visão mais detalhada explorar as informações por intermédio da árvore de decisão do J48.

Neste trabalho foi possível identificar de maneira clara o perfil do docente em instituições de ensino superior com índice de satisfatório a ótimo (3 a 5) no estado do Espírito Santo para os dados do censo 2016 do INEP.

Na análise de dados é possível afirmar com base nos resultados obtidos que as instituições classificadas como Universidades Públicas Federais, que em sua maioria localiza-se na capital do estado enquadram-se no perfil de IGC nota D nota próxima da máxima aferida pelo censo INEP 2016, explicitando o caminho de excelência do ensino superior público do estado do Espírito Santo. O estudo apresenta também bons índices nas instituições particulares do estado, que buscam o reconhecimento e a busca de excelência.

Dos perfis identificados com base na Clusterização e Árvore de decisão:

1. Docentes que atuam nas universidades federais do estado o Espírito Santo são doutores com dedicação exclusiva, localizada na capital, com predominância do sexo feminino com uma faixa etária de 36 a 44 anos em média, de raça declarada branca são IGC D.
2. Instituições privadas com fins lucrativos que se encontram no interior do estado, em sua maioria se enquadram como faculdade, seus docentes são na maioria especialistas, com dedicação parcial, homens brancos, com faixa etária de 25 a 35 anos, são IGC C.

3. No Interior, os docentes que se auto declaram negros estão na faixa 25 a 35 anos com mestrado e doutorado são IGC D, já os com especialização são IGC C.
4. No interior os docentes com titulação de mestre com faixa etária de 36 a 54 anos, estão como IGC C.
5. Os Centros Universitários do Espírito Santo são IGC C.

Com este estudo chega-se à conclusão que a titulação do docente ainda tem um grande peso na classificação do IGC, sobretudo nas instituições privadas e que se encontram fora da capital do estado, pois em sua maioria o perfil encontrado foi o C, e que há um predomínio do sexo feminino nas instituições com IGC satisfatório ou superior na capital do estado. Destacamos também a baixa frequência de Negros e Índios com titulação de mestres.

As informações obtidas com este estudo podem ser de grande valia para políticas de inclusão ou equidade para os órgãos competentes, além de ajudar as instituições a planejar melhor, suas ações rumo a excelência nas avaliações, uma vez conhecido o perfil docente com os melhores índices de IGC do estado do Espírito Santo, fica mais fácil traçar estratégias para alavancar os índices e chegar a pontuação 4 ou 5.

12. REFERÊNCIAS BIBLIOGRÁFICAS

- BATISTA, Fábio Barbosa; BARRETO, Willian de Sousa; GOMES, Geórgia Regina Rodrigues. Utilização de Mineração de Dados para descrição do perfil dos docentes das IES com conceito satisfatório no IGC em 2011 do estado do Rio de Janeiro. Revista de Ciência & Tecnologia. Junho de 2015. Universidade de Iguazu (UNIG)- RJ.
- CARDOSO, Olinda Nogueira Paes; MACHADO, Rosa Teresa Moreira. Knowledge management using data mining: a case study of the Federal University of Lavras. Revista de Administração Pública, v. 42, n. 3, p. 495–528, Jun 2008.
- CAMPOS, Raphael; Árvores de Decisão. <<https://medium.com/machine-learning-beyond-deep-learning/%C3%A1rvores-de-decis%C3%A3o-3f52f6420b69>>. Acessado em :10-12-2018.
- CAMILO, Cássio Oliveira; CARLOS, João. Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas. RT-INF_001-09. p. 29, 2009. Instituto de Informática. Universidade Federal de Goiás. Goiás-GO.
- CASSIANO, Keila Mara. Análise de Séries Temporais Usando Análise Espectral Singular (SSA) e Clusterização de Suas Componentes Baseada em Densidade. Teste (Doutorado em Engenharia Elétrica) – PUC. Rio de Janeiro. 2014.
- CUNHA, M.I.; FERNANDES, C. M.; FORSTER, M. M. Avaliação externa e os Cursos de Graduação: implicações políticas na prática pedagógica e na docência. Perspectiva para o

ensino e à docência universitária frente ao impacto das políticas de avaliação externa no Brasil. In: DIAS SOBRINHO, J, RISTOFF, D.(Orgs.) Avaliação e compromisso público. A educação superior em debate. Florianópolis: Insular, 2003.p. 97-112.

FAYYAD, Usama; PIATETSKY-SHAPIO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. p. 18, 1996. AI Magazine.

GIL, Antônio Carlos. Como elaborar projetos de pesquisa. 4. ed. São Paulo: Atlas S.a, 2002. 176 p.

INEP, Manual do Enade, 2016 – Brasília: Inep, 2016. Disponível em: <http://download.inep.gov.br/educacao_superior/enade/manuais/manual_do_enade_28092016.pdf>. Acessado em: 13-12-2018.

INEP, SINAES 2015 - volume 5, 2015– Brasília: Inep, 2015. Disponível em: <<http://portal.inep.gov.br/documents/186968/484109/SINAES+-+Sistema+Nacional+de+Avalia%C3%A7%C3%A3o+da+Educa%C3%A7%C3%A3o+Superior+Vol+5/ee07ce50-a60f-4614-9aac-01bfe168df5f?version=1.2>>. Acesso em: 12-12-2018.

JAIN, A.K., MURTY, M.N. & FLYNN, P.J. “Data Clustering: A Review”, ACM Computing Surveys, vol. 31, no. 3, pp. 264-323,1999.

MAIMON, Oded. Introduction To Knowledge Discovery In Databases. Israel, 2009. Disponível em: <<http://www.ise.bgu.ac.il/faculty/liorr/hbchap1.pdf>>. Acesso em: 03/12/2018.

MONTEIRO E SILVA, A. B., PORTUGAL, M. S., CECHIN, A. L. (2001). Redes Neurais Artificiais e Análise de Sensibilidade: Uma Aplicação à Demanda de Importações Brasileira. Revecap vol. 5 n. 4.

QUINLAN, J.R. C4.5: programs for machine learning. Sydney, Austrália: Morgan Kaufmann Publishers, 1993. 302 p.

RISTOFF, Dilvo; GIOLO, Jaime. O Sinaes como Sistema. RBPG, Brasília, v. 3, n. 6, p. 193-265, dez. 2006.

SOCZEK, Felipe Cebulski; ORLOVSKI, Regiane. Mineração de Dados: Conceitos e aplicação de algoritmos em uma Base de Dados na área da saúde p. 25, 2014. Semana Acadêmica – Revista Científica.