



## **MINERAÇÃO DE DADOS APLICADA NA BASE DO CADASTRO ÚNICO PARA PROGRAMAS SOCIAIS COM ENFOQUE NOS PERFIS DOS USUÁRIOS CADASTRADOS UTILIZANDO O ALGORITMO J48**

Recebido: 01/08/2018 Aprovado: 29/01/2019

Lívia da Silva Meato<sup>1</sup>  
Valderedo Sedano Fontana<sup>2</sup>  
Ednea Zandonadi Brambila Carletti<sup>3</sup>

### **RESUMO**

O KDD (Knowledge Discovery in Databases), ou processo de extração do conhecimento em banco de dados, compreende técnicas que possibilitam a verificação e extração de padrões e informações vantajosas de uma grande base de dados, identificando semelhanças nos dados e apresentando regras e padrões como resultados. O Cadastro Único para Programas Sociais do Governo Federal (Cadastro Único) é utilizado como instrumento para identificar e caracterizar famílias de baixa renda, possibilitando que a gestão pública tome conhecimento da realidade socioeconômica da população. O artigo objetiva gerar conhecimento a partir do banco de dados do Cadastro Único, que contém dados relativos aos usuários que efetuaram o cadastro desde o início da implantação do mesmo até 31 de outubro do ano de 2018, no município de Cachoeiro de Itapemirim, ES. A base de dados utilizada foi obtida no site de Consulta, Seleção e Extração de Informações do Cadastro Único – CECAD. Para o alcance do objetivo proposto, a base selecionada passou pelo processo de KDD, além de serem empregadas técnicas de Mineração de Dados e utilização do software WEKA. A tarefa aplicada para minerar os dados foi a de Classificação, com o emprego da técnica de árvore de decisão a partir do algoritmo J48. Após minerar os dados, pôde-se observar a influência da renda familiar nos resultados, assim como a predominância de situações de vulnerabilidade social, como a identificação de famílias que não possuem água canalizada, a qual não tem acesso à coleta de lixo, da mesma forma que a constatação de ausência de escoamento sanitário. Almeja-se que as informações obtidas deste artigo possam ser empregadas para auxiliar a gestão municipal nas tomadas de decisão, no que se refere à definição de ações de implementação de políticas públicas, assim como na criação e pactuação de programas e benefícios sociais.

**Palavras-Chave.** Cadastro Único; Mineração de Dados; Programas Sociais; KDD (Knowledge Discovery in Databases); Algoritmo J48.

---

<sup>1</sup> Graduanda em Engenharia de Produção pela Faculdade Multivix Cachoeiro de Itapemirim, Espírito Santo (Brasil).

<sup>2</sup> Mestre em Pesquisa Operacional e Inteligência Computacional. Professor na Multivix Cachoeiro de Itapemirim, Espírito Santo (Brasil). E-mail: [valderedo@Multivix.edu.br](mailto:valderedo@Multivix.edu.br)

<sup>3</sup> Mestre em Ciência da Informação, Professora e Coordenadora de Pesquisa e Extensão da Multivix Cachoeiro de Itapemirim, Espírito Santo (Brasil).

## DATA MINING APPLIED ON THE BASIS OF SINGLE REGISTRATION FOR SOCIAL PROGRAMS FOCUSING ON THE PROFILES OF REGISTERED USERS USING THE J48 ALGORITHM

### ABSTRACT

O KDD (Knowledge Discovery in Databases), or process of extraction of the knowledge in database, comprises techniques that enable the verification and extraction of patterns and advantageous information from a large database, identifying similarities in the data and presenting rules and standards as results. The Federal Government's Single Register for Social Programs (Single Register) is used as an instrument to identify and characterize low income families, making public management aware of the socioeconomic reality of the population. In this article we aim to generate knowledge from the Single Register database, which contains data related to the users who have registered since the beginning of the implementation of the same until October 31, 2018, in the municipality of Cachoeiro de Itapemirim, ES. The database used was obtained on the site of Consultation, Selection and Extraction of Information of the Single Register. To achieve the proposed goal, the selected base passed through the KDD process, in addition to using Data Mining techniques and using WEKA software. The task applied to mining the data was that Classify, with the use of the decision tree technique from the algorithm J48. After mining the data, it was possible to observe the influence of family income on the results, as well as the predominance of situations of social vulnerability, such as the identification of households that do not have piped water, who does not have access to garbage collection, as well as the finding of absence of sanitary outflow. It is hoped that the information obtained from this article can be used to assist municipal management in decision making, in what concerns the definition of actions of implementation of public policies, as well as in the creation and acceptance of programs and social benefit.

**Keywords:** Single Register; Data Mining; Social Programs; KDD (Knowledge Discovery in Databases); J48 algorithm.

## **INTRODUÇÃO**

A motivação deste trabalho considerou o alto valor de investimento do governo federal, estadual e dos municípios em ações e implementação de programas sociais para os usuários cadastrados no Cadastro Único, tendo em vista que o mesmo se trata de um instrumento utilizado para identificar e caracterizar famílias de baixa renda. Essa identificação e caracterização proporciona informações que apresentam ao governo demandas desse público (MINISTÉRIO, 2015). Para isso, é importante que este trabalho aborde não só as características de renda das famílias e do domicílio em que elas residem, mas também dados referentes ao acesso a serviços públicos de água, saneamento e energia elétrica.

Por meio de Decreto da Presidência da República do Brasil, em 2001, foi criado o Formulário de Cadastramento Único para Programas Sociais do Governo Federal, pois neste ano tiveram início as implantações dos programas de transferência de renda como estratégias para o combate à pobreza no Brasil (WORD, 2016).

Para este trabalho foram utilizadas técnicas para minerar dados, tendo em vista que mineração de dados explora um grande volume de dados com o objetivo de identificar padrões que sejam proveitosos e que se tornarão conhecimento.

O objetivo desta pesquisa é aplicar técnicas de mineração de dados na base do Cadastro Único para Programas Sociais, mais especificamente na base dos cadastros e atualizações de cadastros realizados na cidade de Cachoeiro de Itapemirim, ES, utilizando os dados relativos às famílias cadastradas, de modo que seja possível traçar o perfil das mesmas.

Os resultados apresentados neste trabalho, gerados por meio da mineração, podem ser utilizados na tomada de decisão por parte do governo municipal, com o objetivo de aprimorar a implementação de ações de políticas públicas.

## **REFERENCIAL TEÓRICO**

### **CADASTRO ÚNICO PARA PROGRAMAS SOCIAIS**

O Cadastro Único para Programas Sociais do Governo Federal (MINISTÉRIO DO DESENVOLVIMENTO SOCIAL, 2016) é utilizado como instrumento para identificar e

caracterizar famílias de baixa renda, possibilitando que o governo tome conhecimento da realidade socioeconômica da população.

Ressalta-se que o cadastramento não garante a inclusão nos programas sociais disponíveis, tendo em vista que esses seguem legislação e procedimentos específicos amparados por lei, mas é indispensável para que, quando cumprir os demais requisitos, o indivíduo possa ser beneficiário dos mesmos.

O Cadastro Único não se caracteriza por ser somente um banco de informações sobre as famílias público alvo, abrangendo tecnologias, procedimentos e sistemas eletrônicos; mas, sim, como instrumento base de análise para identificar a demanda da população economicamente vulnerável.

### **KNOWLEDGE DISCOVERY IN DATABASES (KDD)**

Com o crescimento exponencial do volume de armazenamento de dados, cresceu também a preocupação quanto à quantidade incompatível de conhecimento adquirido, percebeu-se, então, a que era necessário extrair conhecimento dos dados armazenados.

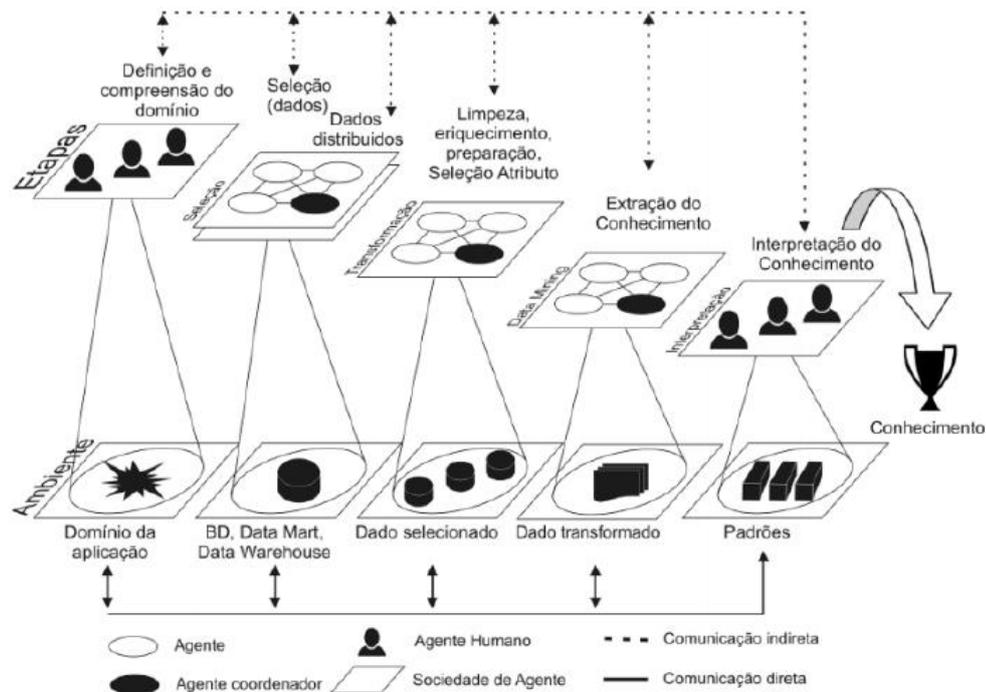
O conhecimento é um importante recurso para auxiliar na tomada de decisões, no desenvolvimento de plano de negócios e na possibilidade de propiciar produtos e serviços de qualidade. Segundo Cardoso e Machado (2008), o processo de controle do conhecimento envolve desde a geração e armazenamento dos dados utilizados para o descobrimento até o ato de identificar e utilizar esse conhecimento.

O crescimento do volume de armazenamento de dados acontece nas mais diversas organizações e suas variadas áreas, contudo, somente o armazenamento de dados não é interessante; torna-se, então, muito importante realizar análise nos bancos de dados, identificar padrões úteis e transformá-los em informação.

Procurar padrões interessantes utilizando dados armazenados não é uma estratégia nova, tendo essa prática já possuído diversos nomes. Segundo Fayyad, Piatetsky-ShapiroSmyth (1996) o termo Knowledge Discovery in Databases (Extração de Conhecimento em Bancos de dados – KDD) só foi adotado em 1989, no 1º *KDD workshop*, que tinha como foco central a conscientização para que houvesse o entendimento de que o objetivo principal da prática era adquirir conhecimento.

Segundo Sassi (2012), utiliza-se de estatística, banco de dados, reconhecimento de padrões e inteligência computacional como métodos para realizar os processos do KDD.

Não existe um padrão correto no processo de extrair conhecimentos a partir de banco de dados (KDD). Assim, esse processo de extração pode apresentar diversos números de etapa (Figura 1).



**Figura 1.** Etapas do KDD.  
Fonte: Cretton (2016)

Mesmo tendo a flexibilidade de se apresentar em diferentes etapas, a estrutura básica do KDD não muda, assim, sua estrutura pode se reduzir em três etapas, sendo elas o pré-processamento, a mineração de dados e o pós-processamento.

As etapas apresentadas na Figura 1 tratam da definição do domínio, da seleção dos dados e da limpeza dos dados, que correspondem ao pré-processamento. A extração do conhecimento representa a mineração de dados e o pós-processamento é representado pela interpretação do conhecimento.

## DEFINIÇÃO E COMPREENSÃO DO DOMÍNIO

Antes de selecionar, limpar e minerar os dados é fundamental definir o assunto, assim como possuir domínio sobre o mesmo. Dessa forma, é importante que nessa fase obtenha-se a assistência de um especialista no assunto definido, que acarreta na extração correta, tal qual

auxiliar na determinação da temática a ser estudada, analisando o conteúdo e desenvolvendo um escopo completo.

## **SELEÇÃO DOS DADOS**

O descobrimento do conhecimento esperado é auxiliado pela correta escolha dos dados, sendo eles conquistados de diversas formas. Santos e Azevedo (2005) afirmam que a base de dados e a estatística são os meios mais comuns de se obter dados para resolver os problemas, contudo, esses dados podem ser obtidos de diversas outras formas.

## **LIMPEZA E TRANSFORMAÇÃO DOS DADOS**

Silva (2007) afirma que a limpeza e transformação dos dados é responsável pelo maior tempo investido no processo do KDD por ser uma tarefa extensa e trabalhosa. É necessário então um método para padronizar os dados, para que haja um formato comum encontrado entre os mesmos.

## **MINERAÇÃO DE DADOS**

Após selecionar, limpar e padronizar os dados, ou seja, o pré-processamento concluído, inicia-se a etapa de minerar os dados. É nessa etapa onde os arquétipos e a gnose são extraídos, sendo extremamente importante a escolha da tarefa e do algoritmo, pois a definição é a causa da qualidade do resultado.

É válido ressaltar que com todo o processo realizado, a mineração de dados pode apresentar diversos resultados, incluindo a chance de resultados que não geram conhecimento. As tarefas de minerar os dados, contidas no software WEKA®, são:

- Classificação: nessa tarefa é necessário definir o atributo central, que é definido na base de dados. Os outros atributos têm sua análise feita baseando-se no atributo escolhido e são separados conforme as semelhanças encontradas. Para a elaboração deste artigo, foi empregada a tarefa de classificação, que utilizou da técnica de árvore

de decisão pelo algoritmo J48, sendo esse responsável por determinar ligações. Após avaliação dos atributos relacionados ao atributo central, dividiu o assunto em problemas menores e tornou suas complexidades menores, aumentando assim a eficiência encontrada na árvore de decisão;

- Regressão: caracterizada pelo emprego de números, possuindo várias variáveis de entrada e apenas uma variável de saída. O software faz um mapeamento onde, tendo como base o atributo de saída, são realizadas análise nas variáveis de entrada;
- Clusterização: gera conjuntos de atributos com base em suas características semelhantes, também conhecida por agrupamento. É uma tarefa que somente apresenta dados de entrada. Segundo Santos e Azevedo (2005), o algoritmo mais utilizado dessa tarefa é o K-means, pois o mesmo é comumente utilizado quando o resultado é desconhecido e não é possível prever qual será;
- Associação: sua análise se baseia nas dependências e seus atributos são separados em dependências e consequências. A associação entre os atributos se dá com base na lógica em que as dependências são condições para as consequências. O algoritmo Apriori é o mais normalmente utilizado, nele são aceitas apenas letras, devendo os dados serem observados no pré-processamento. Esse algoritmo define que a 1ª parte das regras deve ser cumprida antes do algoritmo passar para a 2ª parte e assim sucessivamente.

## INTERPRETAÇÃO DO CONHECIMENTO

Última fase da extração do conhecimento em base de dados, conhecida como interpretação do conhecimento, tendo em vista que é nessa etapa que os resultados obtidos com a mineração de dados passam por análise, faz-se a validação e acontece a interpretação. Assim, todo o processo realizado é analisado, e, quando necessário, fazem-se alterações e o processo recomeça com o propósito de obter melhores resultados. É importante ressaltar que os resultados, para serem considerados vantajosos, devem ser considerados confiáveis, fáceis de compreender e devem apresentar resultados até então desconhecidos.

## SOFTWARE WEKA©

Vianna *et al* (2010), dizem que o software WEKA processa dados de forma rápida e efetiva, além de ser um programa claro e fácil de usar, sendo muito utilizado como ferramenta para minerar dados.

O software possui fácil acesso e interface gráfica amigável, permitindo fácil acesso por meio da aplicação Explorer são a Tarefa de Classificação, Tarefa de Clusterização e Tarefa de Associação.

## METODOLOGIA

O Cadastro Único para Programas Sociais possui uma vasta quantidade de dados em sua base, alguns dados são de domínio público, sendo de fácil acesso à população. Outros dados, onde constam informações pessoais dos usuários cadastrados, são de status sigiloso e somente podem ser acessados por trabalhadores devidamente capacitados e cadastrados.

O software utilizado para a realização das tarefas de mineração de dados deste trabalho foi o WEKA 3.7 (Waikato Environment for Knowledge Analysis), que foi desenvolvido na Nova Zelândia, na Universidade de Waikato.

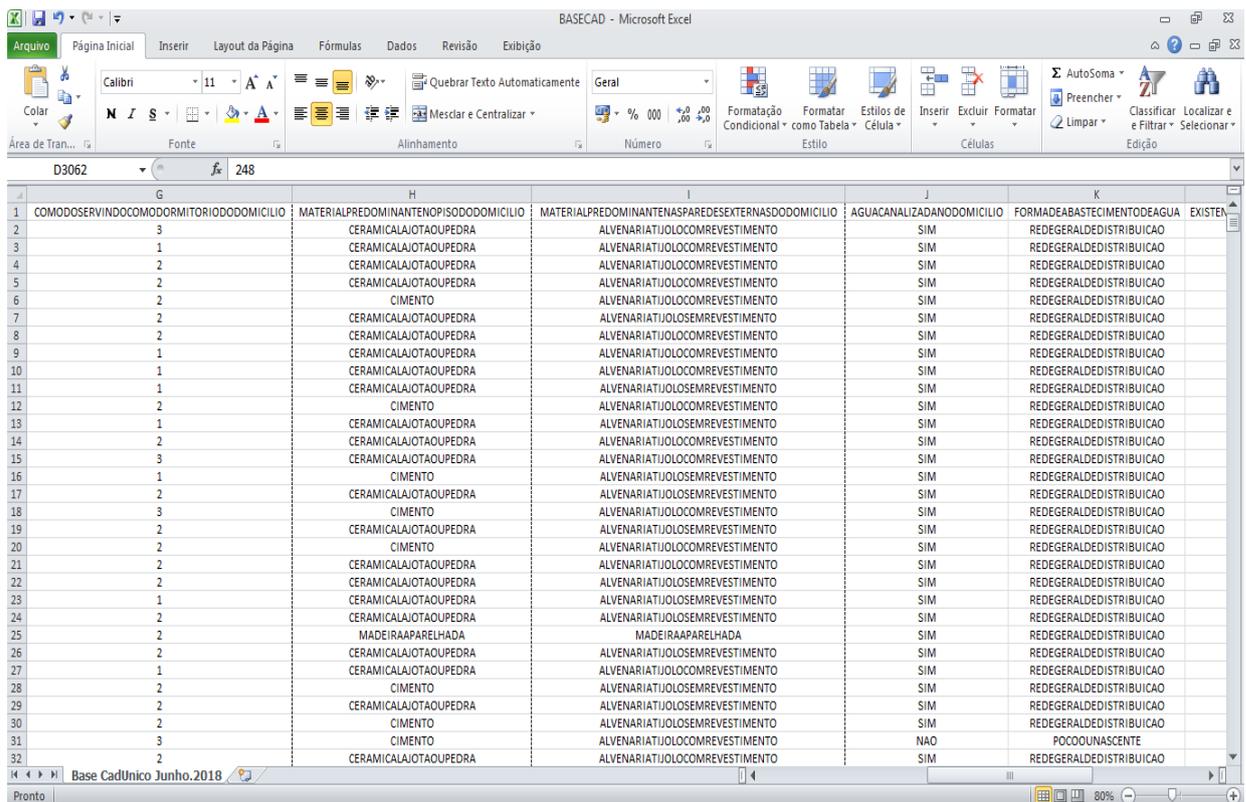
Para Silva (2004), o WEKA é um software intuitivo e com uma interface gráfica amigável que, além de ser gratuito e desenvolvido em Java, o que permite que ele seja utilizado em várias plataformas, também conta com uma grande quantidade de algoritmos, fatores que influenciam no seu alto índice de utilização.

A metodologia foi aplicada na base de dados do Cadastro Único para Programas Sociais, nos cadastros realizados na cidade de Cachoeiro de Itapemirim, ES, com o objetivo de encontrar conhecimentos úteis, que possam ser utilizados para conhecer os perfis de usuários do cadastro. Esse conhecimento pode ser utilizado para a criação de ações de implementação de políticas públicas, assim como ser utilizado como base para maiores pesquisas e desenvolvimento de programas que atendam esse público específico. Foi realizada a preparação da base, a extração da informação e a análise da mesma. O software WEKA foi utilizado para minerar os dados

A base utilizada para a elaboração deste artigo foi adquirida no site de Consulta, Seleção e Extração de Informações do CadÚnico – CECAD, onde os dados somente estão disponíveis para usuários cadastrados. Os registros utilizados provêm da base atualizada em

outubro e contém dados relativos aos usuários que realizaram o Cadastro Único desde o início da implantação do mesmo até 31 de outubro do ano de 2018.

Na base de dados, adquirida em planilha do Excel, é possível encontrar informações como: ano de cadastramento da família, valor da renda familiar per capita, quantidade de cômodos no domicílio, material predominante no piso do domicílio, forma de coleta de lixo, dentre outros, conforme apresentado na Figura 2.



	G	H	I	J	K	L					
	COMODO	SERVINDO	COMODOR	MITOR	ODOD	DOMICILIO					
	MATERIAL	PREDOMINANTE	NO	PISO	DOMICILIO	MATERIAL					
	PREDOMINANTE	NAS	PAR	DE	SE	TERNAS					
	ODOD	DOMICILIO	AGUI	CANALIZADA	NO	DOMICILIO					
	FORMA	DE	ABASTECIMENTO	DE	AGUA	EXISTEN					
1	3	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
2	1	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
3	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
4	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
5	2	CIMENTO			ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
6	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
7	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
8	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
9	1	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
10	1	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
11	1	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
12	2	CIMENTO			ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
13	1	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
14	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
15	3	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
16	1	CIMENTO			ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
17	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
18	3	CIMENTO			ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
19	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
20	2	CIMENTO			ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
21	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
22	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
23	1	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
24	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
25	2	MADEIRA	PAR	ARELHADA	MADEIRA	PAR	ARELHADA	SIM	REDEGERALDE	DISTRIBUICAO	
26	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
27	1	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
28	2	CIMENTO			ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
29	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
30	2	CIMENTO			ALVENARIATU	JULOSEM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	
31	3	CIMENTO			ALVENARIATU	JUOLOCOM	REVESTIMENTO	NAO	POCO	UNASCENTE	
32	2	CERAMICALA/O	TAO	UPEDRA	ALVENARIATU	JUOLOCOM	REVESTIMENTO	SIM	REDEGERALDE	DISTRIBUICAO	

**Figura 2.** Planilha com informações básicas do cadastro único de dados.

**Fonte:** o próprio autor (2018).

No banco de dados utilizado podem ser encontradas informações referentes a 15.462 cadastros de famílias inscritas no Cadastro Único na cidade de Cachoeiro de Itapemirim, ES.

Possuindo diversas informações quanto a questões sociais, habitacionais, estruturais e financeiras, os atributos foram analisados de forma cuidadosa. Foram definidos os dados mais relevantes para essa pesquisa e foi realizada a limpeza da base. Foi eliminado os campos vazios, mantendo na mesma apenas as informações que possuem maior relevância para o alcance dos objetivos propostos.

Após a seleção, limpeza e transformação dos dados deu-se início a etapa de mineração dos dados, que, segundo Cretton e Gomes (2016), tem a finalidade de aplicar técnicas e algoritmos de mineração em banco de dados grandes. Nessa etapa os dados serão explorados e analisados de forma intensa com o objetivo de identificar padrões e extrair informações úteis.

Para a etapa de mineração de dados foi escolhido o software WEKA. Essa ferramenta não “lê” o formato padrão do Excel (sendo esse o formato da base de dados obtida). Tornou-se então necessário transformar o arquivo .xls em formato .CSV (separado por vírgulas) para execução da base de dados na ferramenta escolhida.

Após carregar a base de dados no software, optou-se por usar a tarefa de Classificação, com o algoritmo J48, sendo esse de árvore de decisão.

Foi escolhida a técnica de árvore de decisão, mencionado anteriormente, pois essa possui uma forma simples de observação, tendo em vista que os padrões encontrados pela tarefa de classificação, nesse modelo, são dispostos em formato de árvore, apresentando as ligações de um padrão de forma intuitiva.

Aplicando a tarefa de classificação por meio do algoritmo J48, foi gerado um modelo classificador para a base de dados, juntamente com sua árvore de decisão. Foram extraídas diversas informações e resultados, sendo os mesmos analisados e estruturados, de forma a utilizar os padrões mais pertinentes ao estudo realizado.

## **RESULTADOS E DISCUSSÃO**

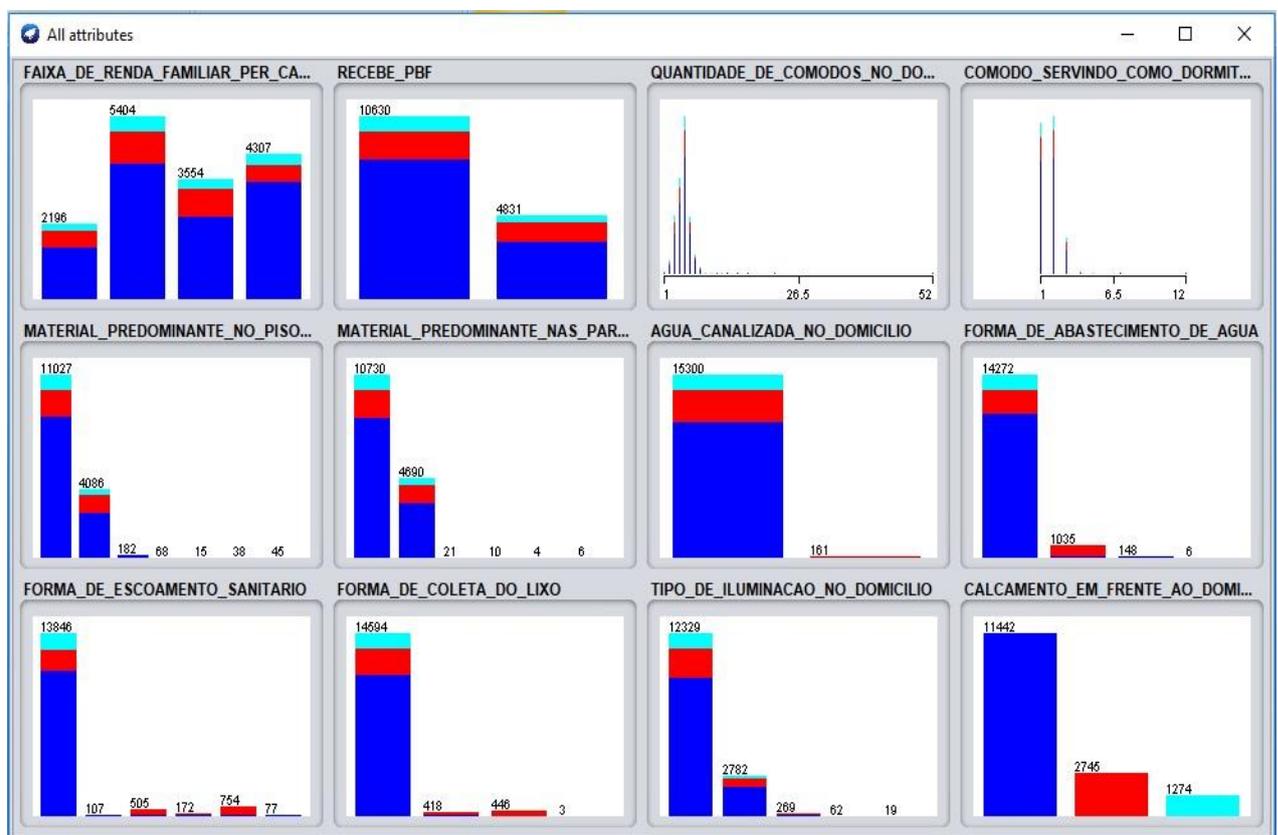
Como os resultados da mineração de dados são apresentados para os usuários, esse é um aspecto básico, pois a usabilidade dos resultados é totalmente dependente. Várias formas de visualização e interfaces estão disponíveis e são usadas nessa etapa (DUNHAM, 2003).

A base utilizada apresenta dados sobre características das famílias e dos domicílios em que elas residem, assim como dados referentes ao acesso a serviços públicos de água, saneamento e energia elétrica. Todos os dados utilizados foram atualizados no mês de outubro de 2018, contudo, apenas 12 atributos foram selecionados por apresentarem resultados relevantes. Segundo Frank (2005), o projeto Weka é baseado em uma coleção organizada de algoritmos de aprendizado de máquina e ferramentas de pré-processamento de dados.

As árvores de decisão apresentadas são provenientes do banco de dados da Consulta, Seleção e Extração de Informações do Cadastro Único - CECAD, da base de dados atualizada em outubro de 2018, que possui uma grande quantidade de dados relativos a usuários

cadastrados no Cadastro Único. Foi realizado um trabalho na base de dados por meio das etapas do KDD, onde ocorreu a seleção dos dados, seu pré-processamento, as transformações necessárias e, por fim, foi realizada a mineração dos dados.

Foram definidos 12 atributos do domicílio para análise a partir da utilização do algoritmo J48: faixa de renda familiar per capita, recebimento do usuário pbf, quantidade de cômodos, número de cômodo servindo como dormitório, descrição do material predominante no piso do domicílio, material predominantes nas paredes externas, descrição do material, água canalizada, forma de abastecimento de água, forma de escoamento sanitário, forma de coleta de lixo, tipo de iluminação, calçamento em frente da casa (Figura 3).

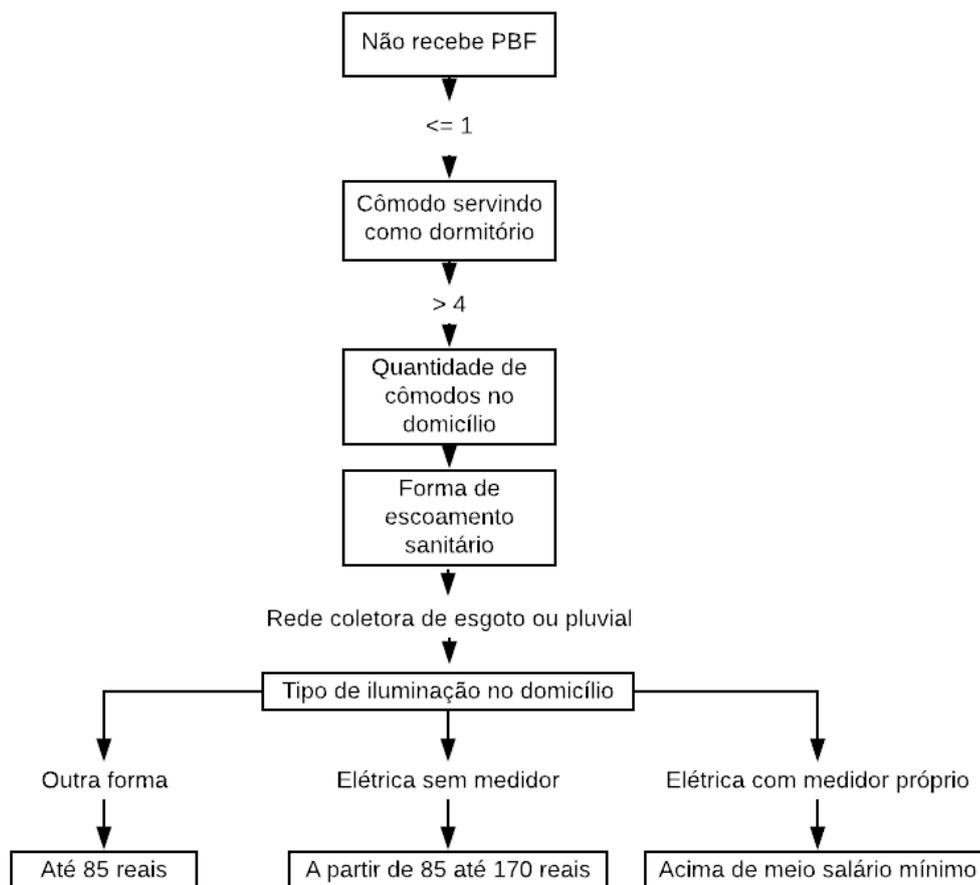


**Figura 3.** Atributos da base de dados do Cadastro Único

**Fonte:** o próprio autor (2018).

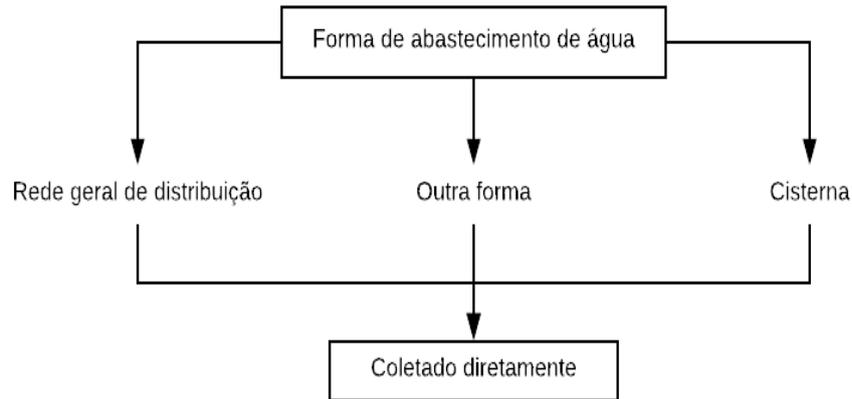
Após o tratamento da base os dados foram minerados com a técnica de classificação utilizando o algoritmo J48, que apresenta regras e árvores de decisão como resultado, sendo esses analisados e apurados, com o objetivo de alcançar resoluções diretas e intuitivas. É possível, então, analisar que nem todos os atributos e variáveis serão apresentados, tendo em vista que os referidos não apresentaram informações importantes para o artigo em questão.

Na análise dos resultados, utilizando como atributo principal à faixa de renda familiar per capita, é possível identificar que os usuários cadastrados que não recebem transferência de renda por meio do Programa Bolsa Família possuem em seus domicílios mais de quatro cômodos - sendo incluso o banheiro – e possuem o número de cômodos servindo como dormitório menor ou igual a um. Ademais, tem seu escoamento sanitário realizado por meio da rede coletora de esgoto ou pluvial, conforme demonstrado na Figura 4.



**Figura 4.** Faixa de renda familiar per capita  
**Fonte:** o próprio autor (2018)

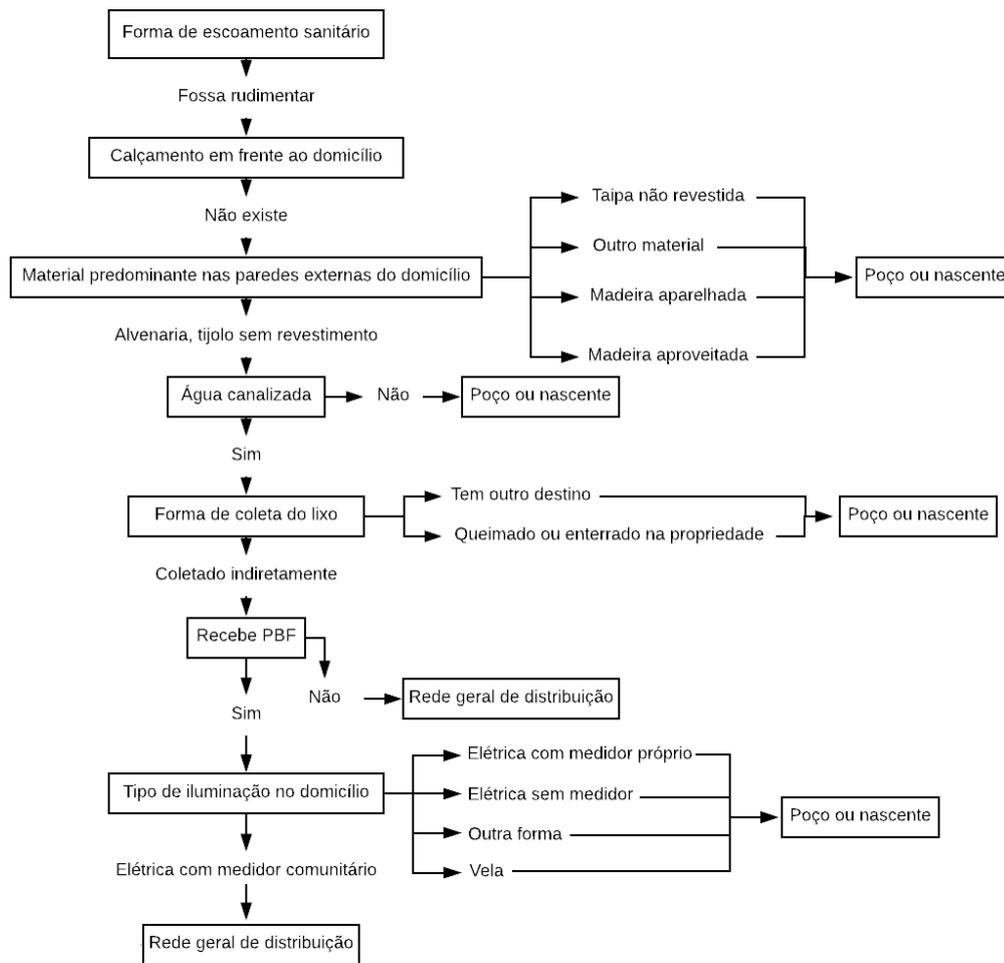
Foi identificado também que, conforme apresentado na Figura 5, os usuários que seu lixo é coletado diretamente têm como forma de abastecimento de água a rede geral de distribuição, cisterna ou outra forma de abastecimento.



**Figura 5.** Forma de coleta do lixo

**Fonte:** o próprio autor (2018)

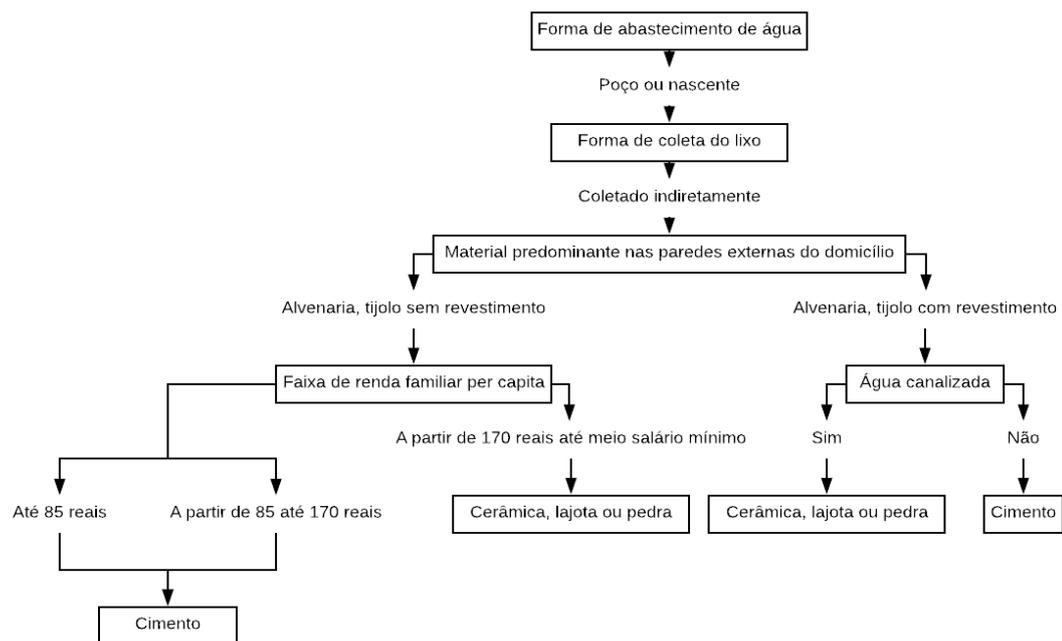
É possível constatar, nos resultados encontrados na árvore que utiliza como atributo principal a forma de abastecimento de água do domicílio, que os usuários que informaram possuir nas paredes externas de seus domicílios materiais como taipa não revestida, outro material, madeira aparelhada ou madeira aproveitada, assim como os que não possuem água canalizada, e, ainda, aqueles que têm seu lixo queimado ou enterrado na propriedade, têm seu abastecimento de água proveniente de poço ou nascente, conforme demonstra a imagem a seguir.



**Figura 6.** Forma de abastecimento de água do domicílio

**Fonte:** o próprio autor (2018)

É possível observar na Figura 7 que a faixa de renda familiar per capita está diretamente ligada ao material predominante no piso do domicílio, assim como a existência de água canalizada ou não.



**Figura 7.** Material predominante no piso do domicílio

**Fonte:** o próprio autor (2018)

A partir da análise, o processo de mineração de dados consiste em extrair dados úteis já presentes em bancos de dados. A base de dados do perfil domiciliar, juntamente com o perfil dos atendidos, é a chave de um problema. Correlacionar essas informações analisando o comportamento desse registro traz uma vantagem estratégica na tomada de decisão, na qual o avaliador pode oferecer gestão somente a um grupo específico de atendidos que realmente necessitam de ajuste (FRANK; WITTENM, 2005).

## CONCLUSÃO

Na utilização da técnica de classificação, por meio do algoritmo J48, foi possível observar que famílias beneficiárias do Programa Bolsa Família apresentam condições de vida menos vulneráveis que famílias com perfil de renda para recebimento do mesmo, mas que não o acessam. Pode-se relacionar alguns conhecimentos na classificação das famílias relacionados à renda.

Na classificação das famílias por sua faixa de renda familiar per capita, foi identificado que as famílias com menor renda não possuem medidor próprio no que tange a iluminação do domicílio. Na árvore de decisão referente à forma de abastecimento de água, é

possível constatar que as famílias que possuem escoamento sanitário por meio de fossa rudimentar, em sua maioria, não têm acesso à rede geral de distribuição de água, precisando recorrer a poços ou nascentes.

Quanto à forma de coleta do lixo, o resultado obtido por meio da utilização do algoritmo J48 mostra uma ligação direta com a forma de abastecimento de água do domicílio.

Como resultados de maior importância, pode se destacar a influência da faixa de renda familiar per capita nos resultados, assim como a predominância de situações de vulnerabilidade social, como a identificação de famílias que não possuem água canalizada, que não tem acesso à coleta de lixo, tal qual a constatação de ausência de escoamento sanitário.

Almeja-se que, com base nos padrões e conhecimentos identificados e expostos, seja possível auxiliar a gestão municipal em sua tomada de decisão, no que se refere à definição de ações de implementação de políticas públicas, assim como na criação e pactuação de programas e benefícios sociais que tenham como objetivo a diminuição da pobreza e o acesso à garantia de direitos, bem como para a redução de vulnerabilidades e riscos sociais.

## REFERÊNCIAS

CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. *Revista de Administração Pública*, Rio de Janeiro, 2008. Disponível em: <<http://www.scielo.br/pdf/rap/v42n3/a04v42n3.pdf>>. Acesso em: 24 nov. 2018.

CRETTON, N. N. *Mineração de dados aplicada na base do ENADE com enfoque na criação de perfis dos estudantes que prestaram o exame utilizando o algoritmo j48*. 2016. 127 f. Dissertação (Mestrado em Pesquisa Operacional e Inteligência Computacional) - Universidade Candido Mendes, Rio de Janeiro, 2016.

CRETTON, N. N.; GOMES, G. R.. *Aplicação de técnicas de mineração de dados na base de dados do ENADE com enfoque nos cursos de medicina*. Acta Biomedica Brasiliensia, 2016. Disponível em: <<http://www.actabiomedica.com.br/index.php/acta/article/view/130/111>>. Acesso em: 01 dez. 2018.

DUHAM, M. H. *Data mining: introductory and advanced topics*. New Jersey: Pearson Education: 2003.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P.. *Advances in knowledge discovery and data mining*. California: AAAI Press, 1996.

FRANK, E.; WITTENM I. H. *Data mining: practical machine learning tools and techniques*. 2nd ed. San Francisco: Morgan Kaufmann Publishers, 2005.

MINISTÉRIO DO DESENVOLVIMENTO SOCIAL. *Cadastro Único para Programas Sociais*, 2015. Disponível em: <<http://mds.gov.br/assuntos/cadastro-unico>>. Acesso em: 30 nov. 2018.

SANTOS, M. F.; AZEVEDO, C. S. *Data mining: descoberta de conhecimento em bases de dados*. São Paulo: FCA, 2005.

SASSI, R. J. *An hybrid architecture for clusters analysis: rough setstheory and self-organizing map artificial neural network*. Pesquisa Operacional, Rio de Janeiro, 2012. Disponível em: <<http://www.scielo.br/pdf/pope/v32n1/aop0512.pdf>>. Acesso em: 24 nov. 2018.

SILVA, A. M. G. *O desempenho escolar via uma abordagem de descoberta de conhecimento em bases de dados*. 2007. 172 f. Dissertação (Mestrado em Sistemas de Informação) - Universidade do Minho, Braga-Portugal, 2007. Disponível em: <<http://repositorium.sdum.uminho.pt/handle/1822/7966>>. Acesso em: 24 nov. 2018.

SILVA, M. P. dos S. *Mineração de Dados - Conceitos, Aplicações e Experimentos com Weka*. 2004. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>>. Acesso em: 20 abril. 2018.

VIANNA, R. C. X. F. et al. *Mineração de dados e características da mortalidade infantil*. Cadernos de Saúde Pública, Rio de Janeiro, 2010. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-311X2010000300011](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2010000300011)>. Acesso em: 25 nov. 2018.

WORLD WITHOUT POVERTY (WWP). *Breve histórico do cadastro unico*. Disponível em: <[https://wpp.org.br/wp-content/uploads/2016/12/breve\\_historico\\_do\\_cadastro\\_unico.pdf](https://wpp.org.br/wp-content/uploads/2016/12/breve_historico_do_cadastro_unico.pdf)>. Acesso em: 01 dez. 2018.