



## Aplicação de Mineração de Dados no Censo da Educação Superior: Estudo de Caso na Base de Docentes de 2014

Bruno Boaventura de Oliveira Lacerda<sup>1</sup>, Manuela Amaral de Araújo<sup>2</sup>,  
Saionara da Silva Araújo<sup>3</sup>, Pablo Freire Matos<sup>4</sup>

<sup>1</sup>Graduando em Sistemas de Informação - IFBA. e-mail: brunoboaventura@gmail.com

<sup>2</sup>Graduanda em Bacharelado em Sistemas de Informação - IFBA. e-mail: mn21922192@gmail.com

<sup>3</sup>Graduanda em Bacharelado em Sistemas de Informação. e-mail: narabdo@gmail.com

<sup>4</sup>Professor de Sistemas de Informação. e-mail: pablofmatos@ifba.edu.br

### INTRODUÇÃO

Com o avanço e redução de custos da tecnologia, é cada vez mais rápido o aumento da quantidade de dados gerados. Alguns desses dados poderiam, se bem analisados, agregar informações importantes para empresas, governos e pessoas. A mineração de dados é uma tarefa muito utilizada para essa finalidade, pois, como parte de um processo de descoberta de conhecimento (KDD), busca gerar informações baseadas em dados gerados por determinado "meio" (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Este trabalho busca demonstrar os resultados obtidos a partir da análise da base de dados do Censo da Educação Superior disponibilizada pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, disponível em <http://portal.inep.gov.br>) com aplicação de algoritmo de regras de associação e com a utilização da ferramenta Weka (2016).

Para o trabalho proposto, a técnica escolhida foi a de associação, indicada quando se deseja descobrir regras que indiquem as relações entre os atributos dos dados informados (SILVA, 2005).

As próximas seções do trabalho tratam dos seguintes pontos: objetivos - são definidos os objetivos geral e específicos da pesquisa; materiais e métodos - traz a ilustração de como a pesquisa foi realizada e o que foi utilizado para alcançar esses objetivos; resultados e discussão - demonstração dos resultados encontrados; conclusão - o que foi possível perceber com o resultado da pesquisa e as possibilidades de utilizar as informações geradas.

### OBJETIVOS

Objetivo Geral: através do conceito sobre o processo de descoberta de conhecimento, aplicar a teoria no desenvolvimento de um projeto prático utilizando a ferramenta Weka e a Associação como técnica de mineração de dados.

Objetivos Específicos:

- Realizar uma busca por uma base de dados recente (a partir de 2010) com uma quantidade razoável de dados;
- Realizar seleção e transformação de atributos para obter regras de associação relevantes;
- Utilizar a ferramenta Weka para analisar a base de dados.

### MATERIAL E MÉTODOS

O processamento e análise dos dados foram realizados utilizando as etapas do processo de KDD proposto por Fayyad, Piatetsky-Shapiro e Smyth (1996), ilustrado na Figura 1:

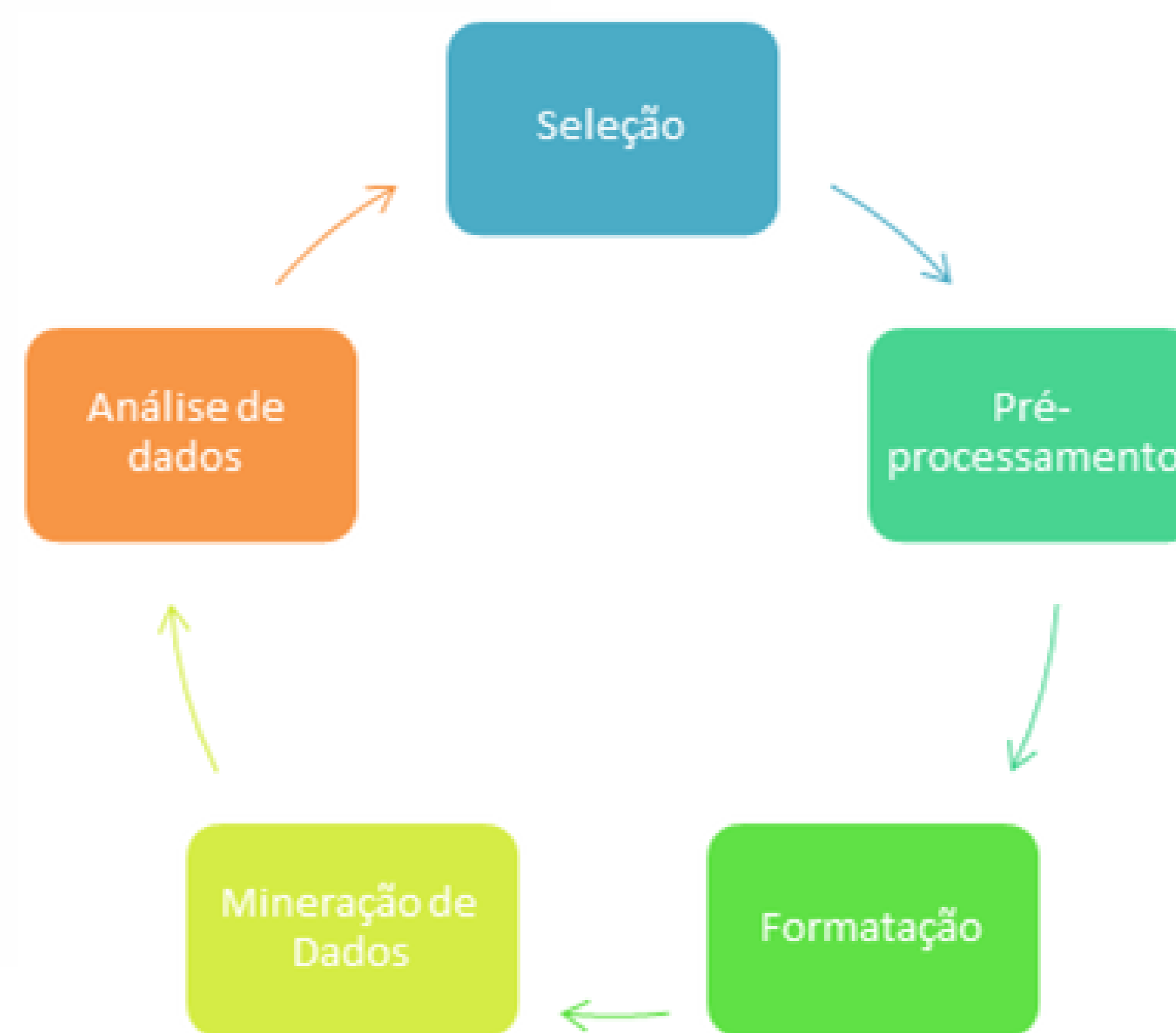


Figura 1. Etapas do trabalho de mineração de dados.

### RESULTADOS E DISCUSSÃO

A Figura 2 representa as regras encontradas após todo o processamento dos dados.

```
Associator output
Instances: 396595
Attributes: 12
NO_IES
DS_CATEGORIA_ADMINISTRATIVA
DS_ORGANIZACAO_ACADEMICA
IN_CAPITAL_IES
DS_SITUACAO_DOCENTE
DS_ESCOLARIDADE_DOCENTE
DS_REGIME_TRABALHO
DS_SEXO_DOCENTE
DS_COR_RACA_DOCENTE
IN_ATU_EXTENSAO
IN_ATU_PESQUISA
IN_BOLSA_PESQUISA
=== Associator model (full training set) ===

Apriori
=====
Minimum support: 0.4 (158638 instances)
Minimum metric <confidence>: 0.99
Number of cycles performed: 12

Generated sets of large itemsets:
Size of set of large itemsets L(1): 9
Size of set of large itemsets L(2): 10
Size of set of large itemsets L(3): 2

Best rules found:
1. IN_ATU_EXTENSAO=N 298463 ==> DS_SITUACAO_DOCENTE=Em exercicio 298463 <conf: (1)> lift: (1.03) lev: (0.03) [994]
2. IN_ATU_PESQUISA=N 277300 ==> DS_SITUACAO_DOCENTE=Em exercicio 277300 <conf: (1)> lift: (1.03) lev: (0.02) [923]
3. IN_ATU_EXTENSAO=N IN_ATU_PESQUISA=N 239989 ==> DS_SITUACAO_DOCENTE=Em exercicio 239989 <conf: (1)> lift: (1.03)
4. DS_SEXO_DOCENTE=Mascúlnno IN_ATU_EXTENSAO=N 166992 ==> DS_SITUACAO_DOCENTE=Em exercicio 166992 <conf: (1)> lift:
```

Figura 2. Tela do Weka com as regras encontradas.

A regra  $IN\_ATU\_EXTENSAO=N \Rightarrow DS\_SITUACAO\_DOCENTE=Em\ exercicio$  mostra que 75,25% dos docentes em exercício não atuam em projetos de extensão. E de acordo com a regra  $IN\_ATU\_PESQUISA=N \Rightarrow DS\_SITUACAO\_DOCENTE=Em\ exercicio$ , a atuação dos docentes em pesquisa é de apenas 30,07%. A terceira regra,  $IN\_ATU\_EXTENSAO=N \wedge IN\_ATU\_PESQUISA=N \Rightarrow DS\_SITUACAO\_DOCENTE=Em\ exercicio$  deixa evidente que 60,50% dos docentes não atuam nem em pesquisa e nem em extensão.

### CONCLUSÃO

A mineração de dados é um valioso recurso para obtenção de conhecimento. Principalmente na atualidade, onde a quantidade de dados aumenta de forma exponencial e a utilização somente de recursos humanos para análise desses dados se torna insuficiente. As ferramentas desenvolvidas, como o Weka, disponibilizam um recurso importante para análise e obtenção de informações valiosas para a tomada de decisão em diversos ramos empresariais e acadêmicos.

Futuramente, se poderia aplicar essa pesquisa em outros dados disponibilizados no próprio site do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), para que os dados sejam utilizados não só para consulta e estatística e sim como fonte de conhecimento para identificar comportamentos comuns que implicam em determinados resultados que, se identificados, poderiam ser um diferencial na tomada de decisões na educação básica e superior. É importante ressaltar que outros atributos poderiam ser utilizados, e outras regras relevantes poderiam ser descobertas, o que pode ser feito em trabalhos futuros.

### AGRADECIMENTOS



### REFERÊNCIAS BIBLIOGRÁFICAS

- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. AAAIIntelligence, p. 37-54, 1996.
- SILVA, G. C. **Mineração de regras de associação aplicada a dados da secretaria municipal de saúde de Londrina - PR**. 94 f. Dissertação de Mestrado - UFRGS, Porto Alegre, 2005.
- WEKA. **Machine Learning Group at the University of Waikato: Downloading and installing Weka**. 2016. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/downloading.html>>. Acesso em 15 jul. 2016.