



UNIVERSIDADE  
FEDERAL DO CEARÁ  
CAMPUS DE RUSSAS

**MINERAÇÃO DE DADOS:**  
**RELATÓRIO DA ANÁLISE DE ASSOCIAÇÃO DE DADOS**  
**UTILIZANDO O ALGORITMO *APRIORI* NA BASE DE DADOS**  
***TRAFFIC ACCIDENT SEVERITY***

**MARIA CLARA DE ABREU SILVA**  
**MARIA ELANNE MENDES RODRIGUES**  
**RUSSAS**  
**2019**

## Sumário

1. Introdução .....	3
2. Análise de Associação de Dados .....	4
Referências .....	8

### 1. Introdução

*Data Mining* consiste sobretudo na análise dos dados após a extração, buscando-se por exemplo levantar as necessidades reais e hipotéticas de cada cliente para realizar campanhas de marketing (DE AMO, 2004).

A Mineração de Dados junta técnicas de máquinas de conhecimentos, reconhecimento de padrões, estatísticas, banco de dados e visualização, para conseguir extrair informações de grandes bases de dados (CABENA *et al.*, 1998).

O conhecimento obtido pelas técnicas de Mineração de Dados é geralmente expresso na forma de regras e padrões.

Devido a sua grande aplicabilidade, as regras de associação encontram-se entre um dos mais importantes tipos de conhecimento que podem ser minerados em bases de dados. Estas regras representam padrões de relacionamento entre itens de uma base de dados.

A mineração de regras de associação consiste em encontrar todas as regras de associação que possuam suporte e confiança maiores ou iguais, respectivamente, a um suporte mínimo (SupMin) e uma confiança mínima (ConfMin), especificados pelo usuário (GONÇALVES, 2007).

A base de dados selecionada para realização deste trabalho foi a *Traffic Accident Severity* que consiste em um conjunto de dados sobre acidentes de trânsito.

Os dados são classificados pelos 14 atributos listados abaixo:

- Unnamed: 0 (Sem nome: 0);
- Strassenklasse (Classe Rua);
- Unfalldatum (Data do Acidente);
- Alter (Idade);
- Unfallklasse (Classe Acidente);
- Lichtverhältnisse (Condições de Iluminação);
- Verletzte Personen (Pessoas Feridas);
- Anzahl Fahrzeuge (Número de Veículos);
- Bodenbeschaffenheit (Qualidade de Solo);
- Geschlecht (Sexo);
- Zeit (Tempo);
- Fahrzeugtyp (Tipo de Veículo);
- Wetterlage (Tempo);
- Monat (Mês).

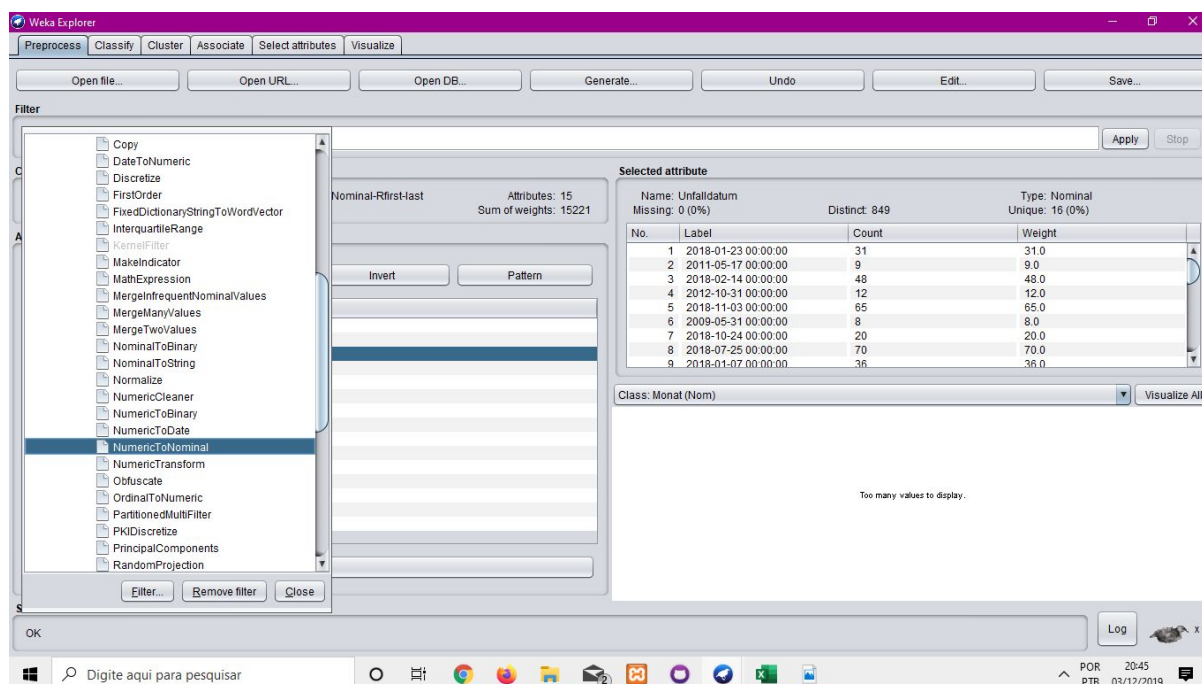
A distribuição dos dados é apresentada em colunas, onde cada coluna representa um dos atributos listados acima.

Existem diversas técnicas de *Data Mining* disponíveis na literatura (CHEN *et al.*, 1996; CHEUNG *et al.*, 1996). Uma das técnicas mais atraentes é a Mineração de Regras de Associação, que tem como destaque o algoritmo Apriori. Ele pode trabalhar com um número grande de atributos, gerando várias alternativas combinatórias entre eles. Esse algoritmo realiza buscas sucessivas em toda a base

de dados, mantendo um ótimo desempenho em termos de tempo de processamento (AGRAWAL & SRIKANT, 1994). Por isso, o escolhemos para realizar a associação retratada neste relatório, cuja atividade foi feita na ferramenta Weka 3.8, optada pela sua facilidade de uso e boa disponibilidade de recursos.

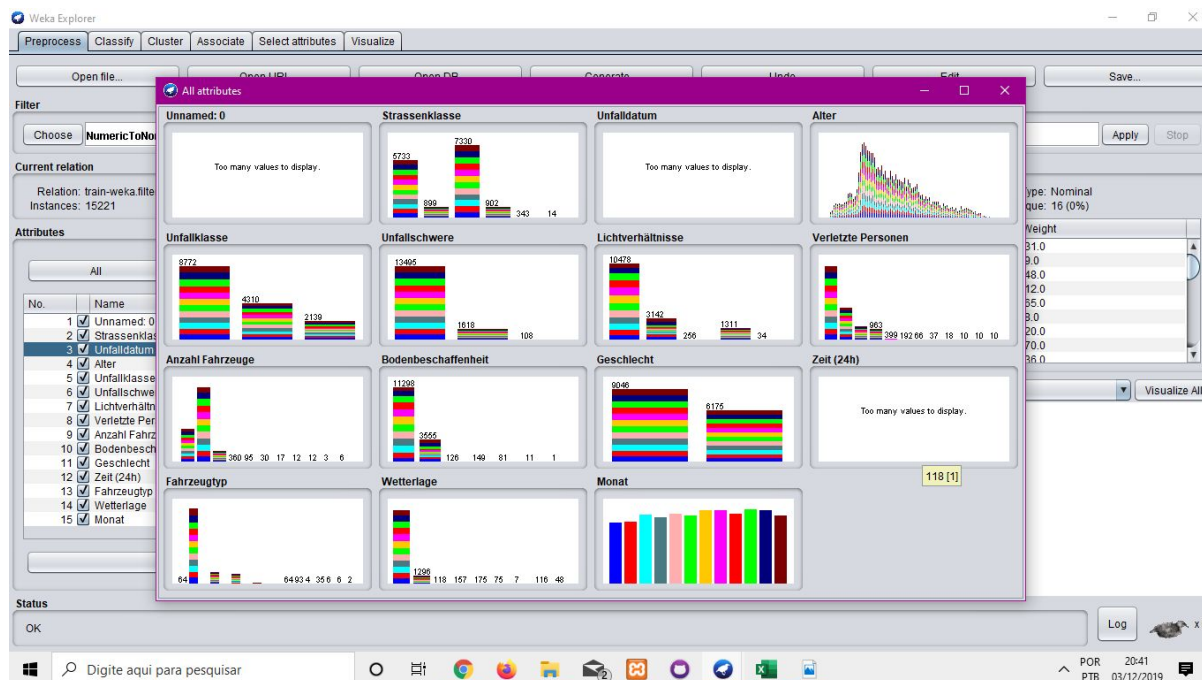
## 2. Análise de Associação de Dados

Ao todo, foram analisadas 15221 instâncias e durante a fase de pré-processamento dos dados, foi realizada a conversão dos atributos para nominal utilizando o método NumericToNominal, como mostra a **Figura 1**.



**Figura 1:** Conversão de atributos no pré-processamento da base de dados na ferramenta.

Na **Figura 2**, podem ser visualizados os gráficos de todos os atributos contidos na base de dados utilizada.



**Figura 2:** Gráficos de todos os atributos da base.

Como resultados gerais, foi obtido como suporte mínimo 0.45 para 6849 instâncias analisadas, uma métrica de confiança de 99% e o algoritmo rodou por 11 vezes, como mostra a **Figura 3**. Como resultado dessa mineração, foram obtidas dez regras, as quais estão dispostas na **Tabela 1** com suas respectivas interpretações.

```
Minimum support: 0.45 (6849 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11
```

**Figura 3:** Resultados gerais.

**Tabela 1:** Regras geradas.

ID	Regra	Interpretação
1	Lichtverhältnisse=Tageslicht: Strassenbeleuchtung vorhanden Bodenbeschaffenheit=trocken 8381 ==>Wetterlage=Gut 8260 <conf:(0.99)> lift:(1.13) lev:(0.06) [975] conv:(8.99)	Com nível de confiança de 99%, se 8381 dos acidentes que ocorreram durante o dia, cuja iluminação era disponível e também ocorreram em um solo seco, então em 8260 desses casos as condições climáticas eram consideradas boas.

2	Unfallschwere=1 Lichtverhältnisse=Tageslicht: Strassenbeleuchtung vorhanden Bodenbeschaffenheit=trocken 7482 ==> Wetterlage=Gut 7370 <conf:(0.99)> lift:(1.13) lev:(0.06) [867] conv:(8.67)	Com nível de confiança de 99%, se 7482 dos acidentes foram de gravidade 1, ocorreram durante o dia, cuja iluminação era disponível, e o solo era seco, então as condições climáticas naquele dia eram consideradas boas para 7370 desses casos.
3	Bodenbeschaffenheit=trocken 11298 ==> Wetterlage=Gut 11128 <conf:(0.98)> lift:(1.13) lev:(0.09) [1308] conv:(8.65)	Com nível de confiança de 98%, se 11298 dos acidentes ocorreram em locais de solo seco, então 11128 destes foram em condições climáticas consideradas boas.
4	Bodenbeschaffenheit=trocken Fahrzeugtyp=Auto 7464 ==> Wetterlage=Gut 7351 <conf:(0.98)> lift:(1.13) lev:(0.06) [863] conv:(8.57)	Com nível de confiança de 98%, se 7464 dos acidentes ocorreram em um local de solo seco e por um carro, então em 7351 desses casos as condições climáticas eram consideradas boas.
5	Anzahl Fahrzeuge=2 Bodenbeschaffenheit=trocken 6999 ==> Wetterlage=Gut 6891 <conf:(0.98)> lift:(1.13) lev:(0.05) [807] conv:(8.4)	Com nível de confiança de 98%, se 6999 dos acidentes envolveram dois veículos e ocorreram em um tipo de solo seco, então em 6891 destes casos as condições climáticas eram consideradas boas.
6	Unfallschwere=1 Bodenbeschaffenheit=trocken 10005 ==> Wetterlage=Gut 9849 <conf:(0.98)> lift:(1.13) lev:(0.08) [1153] conv:(8.34)	Com nível de confiança de 98%, se 10005 dos acidentes possuíam gravidade 1 e ocorreram em solo seco, então as condições climáticas do local eram consideradas boas em 9849 destes casos.
7	Anzahl Fahrzeuge=2 9268 ==> Unfallschwere=1 8504 <conf:(0.92)> lift:(1.03) lev:(0.02) [286] conv:(1.37)	Com nível de confiança de 98%, se 9268 dos acidentes ocorreram com dois veículos ao mesmo tempo, então a gravidade destes foi considerada nível 1 em 8504 dos casos.
8	Anzahl Fahrzeuge=2 Wetterlage=Gut 8071 ==> Unfallschwere=1 7384 <conf:(0.91)> lift:(1.03) lev:(0.01) [228] conv:(1.33)	Com nível de confiança de 91%, se 8071 dos acidentes ocorreram com dois veículos ao mesmo tempo e as condições climáticas do local eram consideradas boas, então em 7384 dos casos a gravidade do acidente foi considerada nível 1.
9	Fahrzeugtyp=Auto 10376 ==> Unfallschwere=1 9463 <conf:(0.91)> lift:(1.03) lev:(0.02) [263] conv:(1.29)	Com nível de confiança de 91%, se 10376 dos acidentes envolveram o veículo carro, então a gravidade do acidente foi considerada de nível 1 em 9463 desses casos.

10	Fahrzeugtyp=Auto Wetterlage=Gut 8890 ==> Unfallschwere=1 8086 <conf:(0.91)> lift:(1.03) lev:(0.01) [204] conv:(1.25)	Com nível de confiança de 91%, se 8890 dos acidentes ocorreram com carro e as condições climáticas eram boas, então a gravidade de 8086 dos casos pode ser considerada de nível 1.

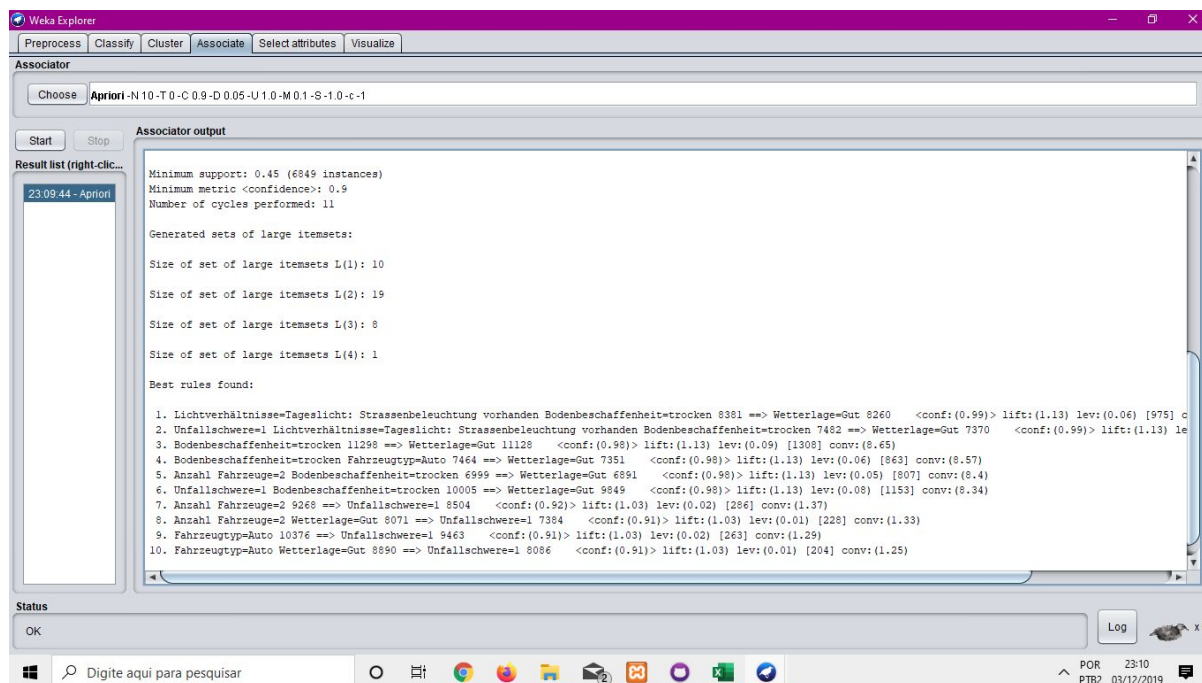


Figura 4: .Visão geral dos resultados.

## Referências

AGRAWAL, R. & SRIKANT, R. **Fast algorithms for mining association rules.** Proc. of the 20th Int'l Conference on Very Large Databases. Santiago, Chile, set. 1994.

CABENA, P; HADJINIAN, P; STADLER, R; JAAPVERHEES; ZANASI, A.  
**Discovering Data Mining: From Concept to Implementation.** Prentice Hall, 1998.

CHEN, M-S., HAN, J. E YU, P. S. **Data mining:** an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering, v. 8, n. 6, p.886-883, 1996.

CHEUNG, D. W., NG , V. T. & FU, A. W. **Efficient mining of association rules in distributed databases.** IEEE Transactions on Knowledge and Data Engineering, v.8, n. 6, p. 911-922, 1996.

DE AMO, S. **Técnicas de mineração de dados.** Jornada de Atualização em Informática, 2004.

GONÇALVES, E. C. **Data Mining de Regras de Associação.** 2007. Disponível em: <<https://www.devmedia.com.br/data-mining-de-regras-de-associacao/6941>>. Acesso em: 03 de dez. de 2019.