

## **Mineração de dados em bases jurídicas: um estudo de caso**

### ***Data mining in legal basis: a case study***

### ***Minería de datos en bases legales: un estudio de caso***

**Talita de Souza Rampão**

Universidade Federal do Paraná (UFPR)  
Brasil

**Denise Fukumi Tsunoda**

Universidade Federal do Paraná (UFPR)  
Brasil

**Submetido em: 21/04/2021**

**Aceito em: 14/06/2021**

**Publicado em: 28/10/2021**

**Licença:**



Autor para correspondência: Talita de Souza Rampão

Email: [talitasouza.qb@gmail.com](mailto:talitasouza.qb@gmail.com)

ORCID: <http://orcid.org/0000-0002-0081-5088>

### **Como citar este artigo:**

TSUNODA, Denise; RAMPÃO, Talita. TCC Mineração de dados em bases jurídicas: um estudo de caso. **REBECIN**, São Paulo, v. 8, edição especial, p. 1-14, 2021. DOI: 10.24208/rebecin.v8i.249

## RESUMO

Apresenta o resultado de um TCC (Trabalho de Conclusão de Curso) desenvolvido no curso de Gestão da Informação referente a aplicação da mineração de dados em uma base de dados jurídica contendo processos cíveis de direito do consumidor. A pesquisa objetivou a aplicação de técnicas de mineração de dados na área jurídica para verificar a existência de padrões de decisões judiciais de acordo com o Estado em que tramitava o processo. No estudo foi realizada a aplicação das tarefas de classificação e associação por meio dos métodos Apriori, PART, *Decision Table*, J48 (C4.5) e REPTree. Os resultados demonstram que é possível identificar padrões de decisões judiciais de acordo com o órgão julgador, tipo de ação e região (Estado) que tramitava o processo. Propõe a análise e continuidade do estudo para verificar a aplicação de técnicas de mineração em outras bases de dados jurídicas, a fim de validar a proposta e comparar as variações nos resultados obtidos.

**Palavras-Chave:** Direito; Gestão da Informação; Descoberta de Conhecimento em Bases de Dados; Mineração de Dados; Tomada de Decisão.

## ABSTRACT

It presents the result of a term paper developed in the Information Management course about the application of data mining in a legal database containing civil lawsuits of consumers. The research aimed to apply data mining techniques in the legal area to verify the existence of patterns or tendencies of judicial decisions according to the State in which the lawsuit was processed. The study performed the application of classification and association tasks through the Apriori, PART, *Decision Table*, J48 (C4.5), and REPTree methods. The result shows that it is possible to predict trends in judicial decisions according to the judging place, type of lawsuit, and region that processed the lawsuit. It proposes the study's analysis and continuity to verify the application of data mining techniques in other legal databases to validate the proposal and compare the variations in the results obtained.

**Keywords:** Law; Information Management; Knowledge Discovery in Database; Data Mining; Decision-Making.

## RESUMEN

Presenta el resultado de un TCC (Course Completion Work) desarrollado en el curso de Gestión de la Información sobre la aplicación de la minería de datos en una base de datos legal que contiene juicios civiles bajo derecho del consumidor. La investigación tuvo como objetivo aplicar técnicas de minería de datos en el ámbito jurídico para verificar la existencia de patrones de decisiones judiciales según el Estado en el que se tramitó el caso. En el estudio, las tareas de clasificación y asociación se aplicaron utilizando los métodos Apriori, PART, Decision Table, J48 (C4.5) y REPTree. Los resultados demuestran que es posible identificar patrones de decisiones judiciales según el órgano de juzgamiento, tipo de acción y región (Estado) que manejó el caso. Propone el análisis y continuidad del estudio para verificar la aplicación de técnicas mineras en otras bases de datos legales, con el fin de validar la propuesta y comparar las variaciones en los resultados obtenidos.

**Palabras llave:** Derecho; Gestión de la información; Descubrimiento de conocimientos en bases de datos; Procesamiento de datos; Toma de decisiones.

## 1 INTRODUÇÃO

A Era da Informação trouxe mudanças no paradigma da sociedade, facilitando o acesso, uso e compartilhamento instantâneo das informações com o auxílio das Tecnologias de Informação e Comunicação (TIC). De acordo com Sidney (2010), essa grande quantidade de dados torna a análise humana onerosa e métodos tradicionais de recuperação de dados, mesmo que sejam sofisticados, não são eficazes para descoberta de conhecimentos “ocultos” em massas de dados.

Para auxiliar na tomada de decisão Caetano (2000) afirma que ferramentas estão sendo utilizadas, de modo a auxiliar na análise dos problemas, bem como transformar informações complexas em informações relevantes, auxiliando na avaliação de resultados. Nesse

contexto, surge como alternativa o KDD (*Knowledge Discovery in Databases*), um processo de descoberta de conhecimento que visa auxiliar na recuperação de informações relevantes a partir de análise de padrões de grandes conjuntos de dados. O KDD conta com as fases de seleção, pré-processamento, transformação, mineração dos dados e interpretação de resultados.

Entre todas as etapas do KDD, a etapa de mineração de dados recebe maior destaque na literatura, considerando que passou a ser vista como um diferencial competitivo, auxiliando os tomadores de decisão a realizarem escolhas estratégicas.

A mineração de dados vem sendo aplicada em diversos segmentos do mercado, apoiando os profissionais nas tomadas de decisão. Castro e Ferrari (2016, p. 17) demonstram alguns exemplos da aplicação: análise e predição de crédito, detecção de fraudes, predição do mercado financeiro, relacionamento com clientes, predição de falência corporativa, entre muitas outras.

No campo de atuação jurídico foi identificado um grande potencial de estudo, considerando que existem grandes volumes de processos tramitando nos tribunais brasileiros, tornando complexo extrair padrões entre as decisões proferidas devido à falta da uniformização processual. Com isso, essa pesquisa descreve a aplicação de técnicas de mineração de dados sobre uma base jurídica cedida por uma organização atuante no ramo, de modo a identificar se existem padrões, conforme o Estado em que tramita o processo.

## **2 KNOWLEDGE DISCOVERY IN DATABASE (KDD)**

O *Knowledge Discovery In Database* consiste no processo de descoberta de padrões pela análise de grandes conjuntos de dados, tendo como principal etapa o processo de mineração, consistindo na execução prática de análise e de algoritmos específicos que, sob limitações de eficiência computacionais aceitáveis, produz uma relação particular de padrões a partir de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Fayyad, Piatetsky-Shapiro e Smyth (1996, p.41) afirmam que o processo de KDD é interativo e iterativo, envolvendo vários passos com muitas decisões tomadas pelo usuário. Os autores consideram o processo de KDD dividido em nove etapas, conforme detalhamento apresentado na sequência:

**Figura 1: Processos de KDD**



Fonte: Free Templates<sup>1</sup> modificado de Fayyad, Piatetsky-Shapiro e Smyth (1996, p. 41)

<sup>1</sup> Your Free Templates. Disponível em: <https://yourfreetemplates.com/free-data-mining-template/>. Acesso em: 07 abr. 2021.

Em resumo, os passos são: (a) conhecimento do domínio da aplicação; (b) criação de um banco de dados alvo por meio da seleção de um conjunto de dados ou dar ênfase para um subconjunto de variáveis ou exemplo de dados nos quais o ‘descobrimento’ será realizado; (c) consiste na limpeza de dados e pré-processamento por meio de operações básicas tais como remover ruídos, coleta de informação, estratégias de tratamento e outros; (d) redução e transformação dos dados; (e) escolha da tarefa de mineração de dados; (f) contempla a avaliação e interpretação do modelo descoberto e possível retorno a algum passo anterior; (g) utilização do descobrimento obtido.

## **2.1 Mineração de Dados**

A mineração de dados (*Data Mining*, em inglês) é um dos principais passos no processo de KDD, tendo sido utilizada para melhorar sistemas de recuperação de informações. A ela corresponde parte da descoberta de conhecimento em bases de dados (KDD), tendo surgido a partir da necessidade de desenvolver ferramentas mais eficientes e escaláveis que pudessem lidar com diversos tipos de dados. (TAN; STEINBACH; KUMAR, 2009, p. 7).

O objetivo da mineração de dados é a extração de conhecimento implícito por meio da descoberta de padrões e regras significativas, a partir de grande quantidade de dados armazenados, de forma automática ou semiautomática, utilizando modelos computacionais construídos para descobrir novos fatos e relacionamentos entre dados, de forma repetida e interativa (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

De acordo com a abordagem de Tan, Steinbach e Kumar (2009, p. 9), existem quatro tarefas centrais da mineração de dados: modelagem

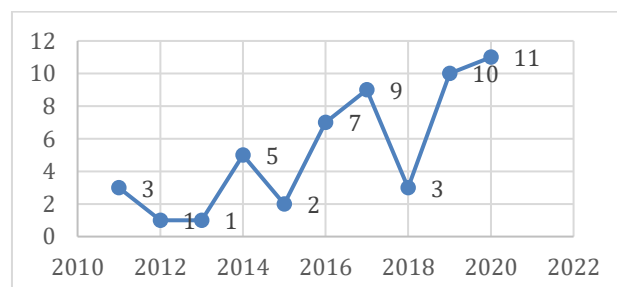
preditiva, análise de associação, detecção de anomalias e análise de agrupamentos.

## 2.2 Mineração de Dados e bases jurídicas

Para que fosse identificada a relevância do tema, mesmo após quatro anos de defesa do referido Trabalho de Conclusão de Curso, foi conduzida uma pesquisa na base Web Of Science (WOS), no dia 07 de abril de 2021, com os descritores ((*data mining OR knowledge discover in database OR text mining*) AND *justice*) em inglês e em português. Não obtendo retorno para os termos em português, passou-se para a análise dos 68 resultados com os descritores em inglês nas ferramentas VosViewer e Biblioshiny (em linguagem R).

Em relação aos anos, nos últimos 10 anos, verifica-se um aumento nas publicações relacionadas ao tema pesquisado, conforme Gráfico 1.

**Gráfico 1: Publicações por ano**

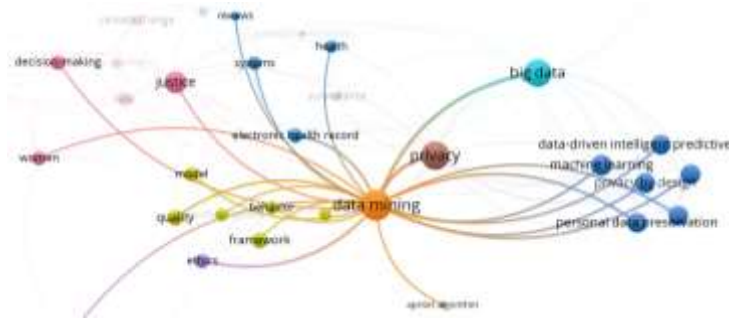


Fonte: Elaborado pelas autoras (2021).

Em relação aos países, dos 68 trabalhos, 26 (38,24%) são dos EUA, 10 (14,71%) da China, 5 (7,35%) da Austrália, 4 (5,88%) da Holanda e 3 (4,41%) da Inglaterra. O Brasil não apresenta publicação relacionada ao tema na base pesquisada.

O mapa de relacionamento das palavras-chaves dos autores permite identificar a relação de “data mining” com “privacy” e também “big data”, conforme Figura 2.

**Figura 2:** Rede de relações das palavras-chaves dos autores



Fonte: Elaborado pelas autoras (2021).

Por meio da análise das palavras-chaves estendidas em nuvem, observa-se novamente a prevalência de alguns termos, dentre os quais destacam-se: privacy, big data, decision making e frameworks, conforme Figura 3.

**Figura 3:** Nuvem de palavras-chaves estendidas



Fonte: Elaborado pelas autoras (2021).

### 3 RESULTADOS E DISCUSSÃO



A pesquisa foi realizada em uma organização privada de Curitiba atuante no segmento jurídico. A base de dados analisada contava com, aproximadamente, mil processos cíveis de direito do consumidor. O atributo meta da análise consistiu no motivo de arquivamento de processos jurídicos com a finalidade de verificar se existe um padrão de decisões judiciais baseado no tipo de ação, região (UF) e órgão julgador que tramita o processo.

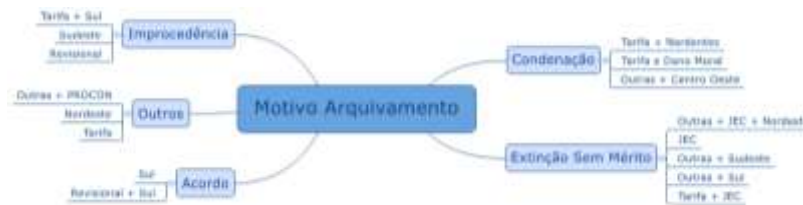
Para proceder a análise da base foram utilizados os métodos de classificação e associação para dar suporte a análise dos dados. Na tarefa de classificação foram utilizadas as heurísticas de regras e árvores, pois apresentam maior facilidade para compreensão dos resultados. Na heurística de regras foi utilizado o algoritmo PART e Decision table, enquanto na heurística de árvores o algoritmo J48 e REPTree. Já na tarefa de associação foi utilizado o algoritmo Apriori, considerando ser o método mais conhecido para mineração de regras de associação.

Na execução do algoritmo Apriori foram obtidas regras com a combinação das variáveis “Ação”, “Região”, “Órgão Julgador” e “Motivo Arquivamento” que permitiram concluir que as ações classificadas como “Outras” tramitando na região sudeste e pelo PROCON apresentaram motivo de arquivamento “Outros”.

Como método de classificação foi utilizado do PART que retorna como os resultado da árvore de decisão resultante da aplicação do método C4.5 em formato de regras. Com base nos resultados obtidos pelo experimento demonstrado na Figura 4 é possível verificar a tendência de realização de acordos em ações revisionais que tramitam na região sul. Também é possível analisar que são comuns condenações em ações de cobrança de tarifa e dano moral, ações de tarifa tramitando na região nordeste e ações “outras” tramitando na região centro-oeste. Dessa

forma, esses seriam os casos mais críticos que deveriam ser analisados pelo escritório de advocacia para conseguir reduzir a quantidade de condenações.

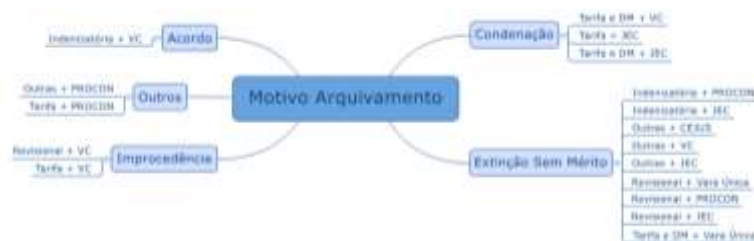
**Figura 4:** Mapa conceitual algoritmo Apriori



Fonte: Elaborado pelas autoras (2016).

Com o algoritmo Decision Table foram obtidas 17 regras que foram sintetizadas no mapa conceitual demonstrado na Figura 5. Com base nos resultados é possível verificar que a classificação dos processos com “Extinção Sem Mérito” ainda são os que apresentam maior dificuldade no reconhecimento de padrões, encontrando quase todos os tipos de ações e órgãos julgadores.

**Figura 5:** Mapa Conceitual algoritmo Decision Table



Fonte: Elaborado pelas autoras (2016).

Com a utilização do C4.5, no Weka J48, um método de árvore de decisão, verificou-se a raiz da árvore como sendo o atributo “Órgão Julgador”, sendo este, portanto, o atributo com maior influência. O

segundo atributo com maior influência corresponde ao “tipo de ação”, seguido de “região”. Para PROCON, Vara Única e Cejusc o resultado foi simplificado, mostrando diretamente o motivo do arquivamento. Já pra VC e JEC existem outros atributos que exercem influência sobre o motivo do arquivamento. Com base no resultado da árvore de decisão foi gerado o mapa conceitual demonstrado na Figura 6 para sintetizar os resultados de acordo com o motivo de arquivamento do processo.

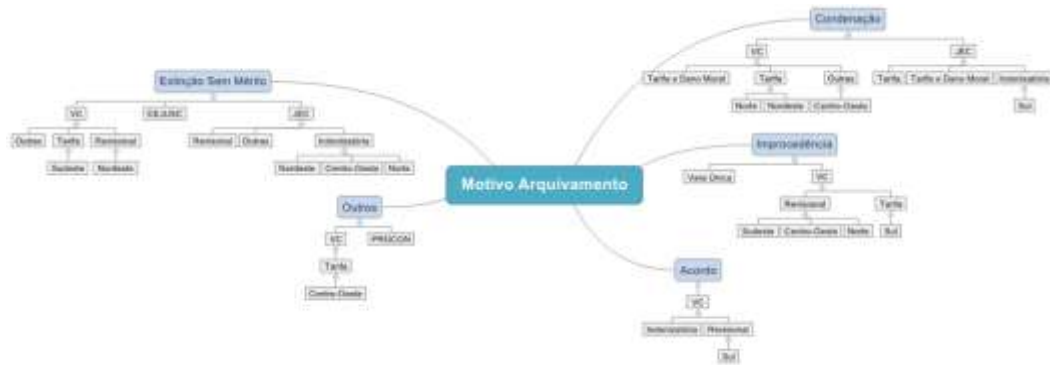
**Figura 6:** Mapa Conceitual algoritmo J48



Fonte: Elaborado pelas autoras (2016).

Na heurística de árvores também foi executado o algoritmo REPTree. A árvore de decisão gerou com raiz o campo “Órgão Julgador”, sendo este, portanto, o atributo com maior influência. O segundo atributo com maior influência corresponde ao tipo de ação, seguido de região. Para PROCON, Vara Única e Cejusc o resultado foi simplificado, mostrando diretamente o motivo do arquivamento. Já pra VC e JEC existem outros atributos que exercem influência sobre o motivo do arquivamento. Com base nos resultados da simulação foi gerado o mapa conceitual demonstrado na Figura 7.

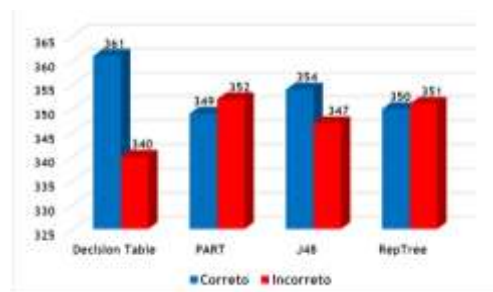
**Figura 7:** Mapa Conceitual algoritmo REPTree



Fonte: Elaborado pelas autoras (2016).

Após a execução de todos os experimentos foi realizada a comparação entre os resultados dos algoritmos com base nas instâncias classificadas correta e incorretamente, conforme demonstrado no Gráfico 2. Analisando os dados é possível identificar que o algoritmo Decision Table realizou uma classificação mais eficiente dos atributos, seguido de J48, REPTree e Part.

**Gráfico 2:** Desempenho de classificação



Fonte: Elaborado pelas autoras (2016).

Os algoritmos J48 e Decision Table atenderam melhor as características da base de dados. Eles conseguiram apresentar resultados satisfatórios para análise, mantendo a acurácia da base de dados e tornando os resultados relevantes para a análise. O J48

apresenta a vantagem de gerar a árvore de decisão que facilita a análise para a tomada de decisão. O Decision Table, por sua vez, gera uma tabela de decisão que também permite analisar condições, contudo, torna a análise mais demorada por não gerar uma representação gráfica. Além disso, o algoritmo gera resultados mais simplificados, tendo em vista que não considera todas as hipóteses possíveis, enquanto na árvore de decisão podem ser analisados todos os caminhos possíveis.

#### **4 CONSIDERAÇÕES FINAIS**

A gestão dos dados e informações tem recebido cada vez mais atenção pelas organizações, considerando que com o adequado tratamento podem-se tornar fonte de vantagem competitiva. Nesse cenário, a mineração de dados tem sido cada vez mais difundida como ferramenta de apoio para a extração de conhecimento nas bases de dados, tendo sido aplicada em diversos ramos de atuação.

No meio jurídico não existem muitos estudos desenvolvidos que auxiliem os profissionais a realizarem a tomada de decisão com base em informações fundamentadas. A pesquisa realizada na WOS aponta, inclusive, que o Brasil não tem publicado no tema pesquisado.

Em relação aos resultados obtidos com a mineração de dados, os algoritmos da tarefa de classificação tiveram resultados mais satisfatórios, considerando que que apresentaram dados mais compreensíveis, permitindo a análise por pessoas que não dominem as técnicas de mineração de dados. Na tarefa de associação, os resultados não foram muito satisfatórios, pois o algoritmo não permite escolher o atributo meta, gerando poucas regras com importância para o enfoque da pesquisa.

Para trabalhos futuros sugere-se a aplicação do estudo em outras bases de dados jurídicas, de forma a validar a proposta e comparar as mudanças nos resultados obtidos. Além disso, é recomendada a aplicação das técnicas em outras áreas do Direito, a fim de verificar se também ocorre a falta de uniformização das decisões jurídicas.

## REFERÊNCIAS

- CAETANO, A. G. L. S. **Sistemas de supervisão de chão-de-fábrica:** uma contribuição para implantação em indústrias de usinagem. 2000. Dissertação (Mestrado em Engenharia) - Escola de Engenharia de São Carlos da Universidade de São Paulo, SP, 2000.
- CASTRO, L. N.; FERRARI, D. G. **Introdução à mineração de dados:** conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996. Disponível em:  
<http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>. Acesso em: 07 maio 2021.
- SIDNEY, C. F. **Aplicação de mineração de dados no banco de dados do zoneamento ecológico econômico de minas gerais.** 2010. 60f. TCC (Graduação em Sistemas de Informação) - Departamento de Ciência da Computação, Universidade Federal de Lavras, Lavras, 2010. Disponível em: <http://goo.gl/zZk0ds>. Acesso em: 07 maio 2021.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining:** mineração de dados. Rio de Janeiro: Ciência Moderna, 2009. 900p.