# Mineração de Dados: Análise de duração de Processos Jurídicos do Estado de São Paulo

**João F. T. da Cunha[1], Wellington F. Silva[1], Anderson F. Talon[1]**

[1]Curso Superior de Tecnologia em Banco de Dados / Fatec Bauru-SP.

{joao.cunha2, wellington.silva18,
anderson.talon}@fatec.sp.gov.br

***Abstract.*** *The backlog and delay in resolving legal cases is a fact known to all. This paper describes the use of data mining techniques in order to perform a detailed analysis of the legal processes of the state of São Paulo. We used the Business Intelligence tool available in SQL Server 2008, where it was found that the neural networks algorithm proved to be the most suitable for the database proposal. It was observed that the amount of data becomes a crucial factor for the choice of algorithm to be used, thus it was concluded that the tax has a higher probability of long term processes have also found that the County Marilia has the most time-consuming processes, followed by Bauru and Santos.*

***Resumo.*** *O acúmulo de processos e a demora na resolução dos casos jurídicos é fato de conhecimento de todos. Este trabalho descreve a utilização da técnica de mineração de dados a fim de se realizar uma análise detalhada dos processos jurídicos do estado de São Paulo. Foi utilizada a ferramenta de Business Intelligence disponível no SQL Server 2008, onde se verificou que o algoritmo de redes neurais mostrou-se o mais adequado para a base de dados proposta. Observou que o volume de dados torna-se um fator crucial para a escolha do algoritmo a ser utilizado, assim, concluí-se que a área tributária possui maior probabilidade de ter processos com longa duração, além disso, verificou-se que a Comarca de Marília tem os processos mais demorados, seguida por Bauru e Santos.*

## 1. Introdução

O sistema judiciário do estado de São Paulo, assim como de todo país, enfrenta um grande problema que não é de hoje: A lentidão processual. O estado é um dos casos mais graves de morosidade, onde milhares de processos acumulam-se aguardando um desfecho. O Tribunal de Justiça conta com uma quantidade de processos acumulados que parecem intermináveis, de acordo com o relatório Justiça em Números, do Conselho Nacional de Justiça, existem mais de 83 milhões de processos em tramitação no país. Em se tratando do estado de São Paulo, esse total chega a mais de 21 milhões, [Costa 2012]. O problema não está limitado apenas a uma determinada instância ou seção, parece ser algo generalizado.

De acordo com Canário (2012) o estado se São Paulo possui cerca de 3 milhões de sentenças por ano, mas recebe 5 milhões de processos, isto gera um déficit de 2 milhões de ações por ano. Ainda segundo o mesmo autor, o estado Paulista possui 2021

juízes e cada um profere 8 sentenças por dia em média. Para atender a atual demanda de processos seriam necessários mais 1092 juízes e que, cada um proferisse 11 sentenças por dia em média.

Segundo pesquisa do Conselho Nacional de Justiça, o Tribunal de Justiça do estado de São Paulo possui um acervo com mais de 600.000 recursos, ou seja, processos com sentença proferida, mas que sofrem uma forma de análise sobre a decisão proferida a fim de se reformá-la, modificá-la ou integrá-la. A pesquisa aponta também que existem 56 desembargadores com mais de 3000 recursos aguardando julgamento. Em fevereiro de 2011 eram 47.782 processos pendentes aguardando julgamento, e a seção considerada mais crítica foi a de direito privado que acumulava mais de 34.000 processos parados, de acordo com Costa (2012).

A tecnologia pode ser uma grande aliada na identificação e busca de soluções para esses problemas. O intuito deste trabalho é utilizar a técnica de mineração de dados para se realizar uma análise detalhada dos dados de processos. A técnica de mineração de dados, no qual faz parte de uma das etapas da descoberta de conhecimento em banco de dados, foi empregada a fim de se procurar por associações que sejam relevantes no auxilio da busca por soluções, apontando os casos que mais demandaram registros, seja pela gravidade do caso ou pela repetição de equívocos de informações.

Assim, propõem-se com este trabalho, uma forma de analisar uma quantidade significativa de processos encerrados com suas datas de distribuições e encerramentos, além das respectivas comarcas, áreas, ações, foros e decisões aos quais estes processos estão sujeitos. Os dados foram modelados da melhor forma possível para que seja feita a descoberta de conhecimento, visando identificar os tipos de processos com maior duração e quais locais encontram-se estes processos.

## 2. *Data Mining*

Uma das técnicas que ultimamente vem ganhando cada vez mais adeptos é a técnica *Data Mining*. Segundo Oliveira (2012), o respeitado instituto de pesquisa *Gartner Group* afirma que as ferramentas de *Data Mining* serão umas das cinco mais importantes tecnologias do século XXI, colocando-a na lista de prioridades dos CIOs (*Chief Information Officer* – Chefe Oficial de Informação) da América Latina, entretanto, poucas instituições são capazes de colocar o modelo em prática de forma eficiente e estruturada.

Mesmo com a popularização do *Data Mining* e as mais variadas tecnologias para extrair informações, a definição do termo pode ser encontrada das mais diversas formas. Sendo assim, algumas definições serão apresentadas para se ter uma ideia sobre o termo referenciado.

*Data Mining* (ou mineração de dados) é o processo de extrair informação válida, previamente desconhecida e de máxima abrangência a partir de grandes bases de dados, usando-as para efetuar decisões cruciais. Pode ser considerada uma forma de descobrimento de conhecimento em bancos de dados (KDD – *Knowledge Discovery in Databases*), área de pesquisa de bastante evidência no momento, envolvendo Inteligência Artificial e Banco de Dados. (CAMPOS; ROCHA FILHO, 1999).

Conforme Harrison (1998), o *Data Mining*, do modo como é usado o termo, pode ser considerado a exploração e análise, por meios automáticos ou

semiautomáticos, de grandes quantidades de dados para descobrir modelos e regras significativas.

A premissa do *Data Mining* é uma argumentação ativa, isto é, em vez do usuário definir o problema, selecionar dados e ferramentas para que se realize a analise, as técnicas e ferramentas do *Data Mining* pesquisam automaticamente estes dados a procura de anomalias e possíveis relacionamentos, identificando assim problemas que não tinham sido identificados pelo usuário. Em outras palavras, as ferramentas de *Data Mining* analisam os dados, descobrem problemas ou oportunidades escondidas nos relacionamentos dos dados, e então diagnosticam o comportamento dos negócios.

De acordo com Berry e Linoff (1997), o objetivo do *Data Mining* é descobrir o conhecimento, extraí-lo implicitamente sem que seja necessário conhecer a estrutura das informações do banco de dados sobre ele aplicado. Este processo é denominado de *Knowledge Discovery in Databases* (Descoberta de conhecimento em base de dados – KDD), o termo KDD refere-se ao processo global de descobrimento de conhecimento útil em bases de dados. *Data Mining* é um passo particular neste processo-aplicação de algoritmos específicos para extrair padrões (modelos) de dados. Os passos adicionais no processo KDD, como: preparação de dados, seleção de dados, limpeza de dados, incorporação de conhecimento anterior apropriado e interpretação formal dos resultados de mineração assegura aquele conhecimento útil que é derivado dos dados. A aplicação cega de métodos de *Data Mining* pode ser uma atividade perigosa que conduz a descoberta de padrões sem sentido.

### 2.2. Redes Neurais Artificiais

Procura imitar as conexões dos neurônios naturais. Recebem a informação e essa passa por várias conexões que aprendem com treinamento e são capazes de retornar dados mais precisos. Provavelmente a técnica mais utilizada para *Data Mining*.

Esta técnica possui algumas desvantagens. Ávila (1998) cita que o processo de aprendizagem pode ser muito lento se compararmos com sistemas de aprendizado simbólico, e o conhecimento gerado não esta representado na forma de regras e padrões e sim implicitamente nas conexões da rede. Podem funcionar melhor quando não haverá informação adicional.

## 3. A Ferramenta de Análise

Para análise dos dados será utilizada a ferramenta *SQL Server analysis services*, que está disponível como uma ferramenta de *business intelligence* da Microsoft. A ferramenta possui sete algoritmos de mineração de dados, abordando todas as categorias, exceto a análise de desvio. O processo começa com a definição da estrutura a ser analisada, que pode ser um banco de dados relacional ou cubo multidimensional. O processamento de um modelo começa pelo treinamento, que nada mais é que recuperar uma parte dos dados e fazer a análise nessa parte para posteriormente usar o resultado desse treinamento e analisar o restante dos dados. É possível configurar a porcentagem de dados a serem usados no treinamento (HOTEK, 2010).

## 4. Descrição dos Experimentos

A extração dos dados foi feita no site do tribunal de justiça do estado de São Paulo, os dados estão disponíveis no portal do Tribunal de Justiça do Estado de São Paulo[1]. A consulta foi realizada escolhendo-se nomes aleatórios de advogados no qual se extraiu dados que são apresentados em forma de tabelas. Estas tabelas foram copiadas para uma planilha e transformadas em arquivos próprios para leitura no *Microsoft Excel* (arquivos com extensão .xls). Usando o recurso de *Linked Servers* (ferramenta que cria um link entre um arquivo .xls e uma instância de banco de dados, possibilitando a comunicação entre os dois) disponível no SqlServer 2008, foi criado um link com esse arquivo .xls para que fosse possível criar a base de dados. A base de dados tem um total de 5740 registros divididos por 7 comarcas.

O arquivo xls é formado pelas seguintes colunas: Número processo, Ação, área, comarca, data distribuição, data encerramento, foro e decisão.

Com o intuito de buscar uma padronização de procedimentos, o Tribunal de Justiça do Estado de São Paulo dividiu o estado em dez áreas denominadas regiões administrativas judiciárias. Cada região administrativa agrupa certo número de circunscrições judiciárias contíguas e tem como sede a comarca que lhe dá o nome (com exceção da Região da Grande São Paulo – 1ª Região). As demais são: Araçatuba (2ª Região Administrativa Judiciária), Bauru (3ª), Campinas (4ª), Presidente Prudente (5ª), Ribeirão Preto (6ª), São José do Rio Preto (7ª), São José dos Campos (8ª), Santos (9ª) e Sorocaba (10ª) (EM BUSCA, 2012).

Neste trabalho, optou-se por algumas Regiões Administrativas Judiciárias (RAs), respeitando as cidades que estão disponíveis para consulta no portal do TJSP. Assim, as cidades ou comarcas definidas para o trabalho foram: Bauru (3ª RA), Campinas (4ª RA), Marília (5ª RA), São José do Rio Preto (7ª RA), Santos (9ª RA), Sorocaba (10ª RA) e São José dos Campos (8ª RA). Foram escolhidos nomes aleatórios de advogados e foram consultados 820 processos por comarca. Este número foi restringido devido ao tempo de execução do trabalho, pelo simples fato da consulta aos processos ter sido realizada de forma manual. A base de dados foi montada com apenas uma tabela e sua estrutura é detalhada na Tabela 1.

Tabela 1: Estrutura da tabela.

| NOME CAMPO | TIPO DE DADOS | VERIFICAÇÃO (*COLLATION*) |
|---|---|---|
| NUMERO | *INTEGER* | *NULL* |
| ACAO | *VARCHAR* | *Latin1_General_CI_AI* |
| AREA | *VARCHAR* | *Latin1_General_CI_AI* |
| COMARCA | *VARCHAR* | *Latin1_General_CI_AI* |
| DATADISTRIBUICAO | *DATETIME* | *NULL* |
| DTENCERRAMENTO | *DATETIME* | *NULL* |
| FORO | *VARCHAR* | *Latin1_General_CI_AI* |
| DECISAO | *VARCHAR* | *Latin1_General_CI_AI* |

Fonte: autores

[1]Os dados são disponíveis através de consulta simples que pode ser obtida no endereço: http://www.tjsp.jus.br/PortalTJ3/Paginas/Pesquisas/Primeira_Instancia/Interior_Litoral_Civel/Por_comarca_interior_litoral_civel.aspx

O campo número é fictício e representa o número do processo. O segundo refere-se a que tipo de ação é o processo, e na referida base tem-se 28 tipos diferentes. O terceiro representa a área do processo base, podendo ser: trabalhista, tributária e cível. O quarto campo representa a cidade em que processo está registrado. O campo data distribuição informa quando o processo teve sua primeira publicação e, o campo data de encerramento é utilizado em conjunto com a data de distribuição para calcular a duração o processo. Em uma cidade pode haver vários foros onde o processo foi registrado, por isso este campo foi considerado no trabalho. E por último a decisão que foi proferida em relação ao processo, que pode ser: procedência da ação, diligência cumprida, extinto sem julgamento do mérito, acordo, encerramento procon, parcial procedência da ação.

Para trabalhos futuros será desenvolvida uma ferramenta capaz de fazer a coleta dos dados de uma forma mais eficiente, disponibilizando assim uma maior quantidade de dados para análise e possivelmente um resultado mais consistente.

## 5. Resultados

Utilizando-se da linguagem SQL (*Structed Query Language* – Linguagem Estruturada de Consulta), realizou-se uma consulta simples, onde se verificou que a duração média de um processo nessas cidades é de 622 dias ou 1 ano 8 meses e 17 dias. A consulta utilizou-se de comandos simples, disponíveis na versão *Transact-SQL*, que compõe a ferramenta *SqlServer 2008*.

```
SELECT AVG(DATEDIFF(DAY, DATADISTRIBUICAO, DTENCERRAMENTO)) FROM PROCESSOS.
```

Com outra consulta foi possível identificar o processo com maior duração, que foi de 6531 dias ou 17 anos 10 meses e 26 dias.

```
SELECT MAX(DATEDIFF(DAY, DATADISTRIBUICAO, DTENCERRAMENTO)) FROM PROCESSOS.
```

Para análise dos dados foi criado um novo projeto no *SqlServer Business Intelligence Development Studio* e, utilizado dois algoritmos: Redes neurais e *Naives* Bayes, pode-se testar os dados disponíveis. Os algoritmos foram testados em computador do tipo *personal computer* com processador dual core com 3 gigabytes de memória RAM. A seguir será detalhado o resultado.

A ferramenta utilizada apresenta diversos resultados possíveis, porém, como a ideia é identificar os processos mais demorados só serão considerados os intervalos que apresentarem a maior duração de tempo, medidos em dias.

Foram realizados vários testes com o algoritmo de *Naives Bayes*, mas pela forma como está montada a base de dados não foram obtidos resultados satisfatórios, sendo assim, ficou decidido que os testes seriam efeitos apenas com o algoritmo de redes neurais, pois o mesmo apresentou resultados mais significativos.

Utilizando o algoritmo de redes neurais foi criado um teste com o campo número sendo utilizado como chave, o campo duração selecionado como predicado e o campo comarca como saída, o percentual de treinamento foi de 50%, e o algoritmo demorou 1

segundo[2] para fazer a análise. O resultado apresentado foi a probabilidade das comarcas terem seus processos com duração entre 1186 dias (3 anos e 3 meses) e 2930 dias (8 anos). A comarca de Marília possui 61,81% de chances, seguida pela comarca de Bauru com 54,67%. Em seguida aparecem as comarcas de Santos com 53,50%, São José do Rio Preto com 43,86% e Sorocaba com 35,68% de chances. Por fim, aprecem as comarcas de Campinas com 33,03% e São José dos Campos com apenas 24,70% de chances.

Outro resultado foi a probabilidade de um processo ter duração entre 680 dias (1 ano e 8 meses) e 1186 dias (3 anos e 3 meses). Neste resultado, a comarca de Sorocaba apresenta maior probabilidade com 25,52%, seguida pelas comarcas de Campinas (25,38%), São José do Rio Preto (25,19%), São José dos Campos (24,02%), Santos (23,50%), Bauru (23,22%) e por fim, Marília, com 21,11%. Observa-se neste resultado que houve pouca variação dos valores.

Utilizando o campo número como chave, o campo decisão como predicado e o campo duração como saída, e definindo os mesmos 50% dos dados como treinamento, o algoritmo demorou 2 segundos para executar sendo foi possível observar que: a probabilidade de um processo com decisões do tipo diligência cumprida com duração entre 994 dias (2 anos e 8 meses) e 2319 dias (6 anos e 4 meses) foi de 82,27%, e que decisões do tipo parcial procedência da ação apresentaram probabilidade de 51,22%, ao passo que decisões do tipo procedência da ação apresentaram 42,75% de probabilidade. O tipo encerramento procon apresentou 39,50%. Já o tipo extinto sem julgamento do mérito apresentou 25,03% de probabilidade e por fim, o tipo acordo ficou com 23,85%.

Foram realizados testes para identificar a área com processos mais demorados, o campo número foi colocado como chave, o campo área como saída e o campo duração como predicado, o algoritmo demorou 17 segundos para rodar e encontrou-se o seguinte resultado: a área tributária possui 88,32% de probabilidade de ter processos com duração entre 1032 dias (2 anos e 10 meses) e 2014 dias (5 anos e 6 meses), a área trabalhista tem 39,38% de probabilidade de processos com a mesma duração e, a área cível apresentou 34,33% de probabilidade.

Como a Comarca de Marília apresentou a maior duração nos processos, foram feitos testes para identificar nesta comarca a duração dos tipos de ação, assim, os seguintes resultados foram obtidos: Para duração entre 1098 dias (3 anos) e 3868 dias (10 anos de 7 meses) a probabilidade para o tipo ação declaratória foi de 70,21%, seguido pelos tipos de ação: reclamação trabalhista (54,79%), condenatória (47,99%), administrativa (44,19%), cobrança (27,83%), conhecimento (27,80%), notificação (25,48%), cautelar (24,32%), ordinária (23,08%), auditar (22,90%), mandado de segurança (10,89%), revisional (8,97%), e consignatória (3,54%).

## 6. Considerações Finais

O tempo médio de duração de um processo é de quase 2 anos (622 dias), e é possível que um processo possa durar mais de 17 anos (6531 dias).

---

[2] A ferramenta disponibiliza o tempo de execução apenas com valores aproximados, não sendo possível visualizar com precisão a sua realização.

Foi possível observar que a área tributária possui maior probabilidade de contar com processos de longa duração. Dividindo o resultado por Comarca, verificou-se que a Comarca de Marília possui os processos mais demorados, seguida por Bauru e Santos.

Analisando a Comarca de Marília também foi constatado que a ação do tipo declaratória tem maiores chances de apresentar uma longa duração, seguida pela a ação condenatória e ação administrativa.

É possível que nessas Comarcas seja considerada a necessidade de alocar mais juízes para atenderem a demanda de processos. No caso de Marília, especialmente para atenderem ação declaratória, condenatória e administrativa. Também é possível uma maior alocação de juízes a fim de atenderem a processos da área tributária.

**7. Referências Bibliográficas**


ÁVILA, B.C. 1998 **Data Mining**. Dissertação (Mestrado em informática Aplicada) – Pontifícia Universidade Católica do Paraná. Curitiba.

BERRY, Michael J. A. e Linoff, G. 1997 **Data Mining techniques**. USA : Wiley Computer Publishing.

CAMPOS, M. L. e ROCHA FILHO, A. V. 2005 **Data warehouse**. Disponível em: <http://genesis.nce.ufrj.br/dataware/tutorial/indice.html>.Acesso em: 03/setembro/2012.

CANÁRIO, P. 2012 **Corregedoria do CNJ Começa Inspeção no TJ-SP**. Disponível em: <http://www.conjur.com.br/2012-ago-06/cnj-comeca-inspecao-tj-sp-foco-atrasos-corrupcao>. Acesso em: 23/outubro/2012.

COSTA, M. **As sequelas criadas pela lentidão da Justiça**. Disponível em <http://www1.folha.uol.com.br/fsp/opiniao/50847-as-sequelas-criadas-pela-lentidao-da-justica.shtml>. Acesso em: 03/setembro/2012.

**EM BUSCA de Padronização de Procedimentos, Judiciário Divide Estado em Regiões Administrativas** 2012 Diário da Justiça Eletrônico. Ano V. 1190ª edição. Disponível em: <http://www.tjsp.jus.br/Handlers/FileFetch.ashx?id_arquivo=40204>. Acesso em: 25/março/2013.

HARRISON, T. H. 1998 **Intranet data warehouse**. São Paulo : Berkeley Brasil.

HOTEK, M. 2010 **Microsoft Sql-Server 2008: Passa à Passo**. Editora Bookman.

OLIVEIRA, D. 2012 **Data Mining ganha espaço na estratégia empresarial**el em: <http://computerworld.uol.com.br/tecnologia/2012/03/16/data-mining-ganha-espaco-na-estrategia-empresarial>. Acesso em: 03/Setembro/2012.

PICHILIANI, M. **Data Mining na Prática: Classificação Bayesiana**. Disponível em: <http://imasters.com.br/artigo/4926/sql-server/data-mining-na-pratica-classificacao-bayesiana>. Acesso em 03/setembro/2012.